

09/327,984

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:50:09 ; Search time 120.01 Seconds
(without alignments)
357.991 Million cell updates/sec

Title: US-09-724-296-38
Perfect score: 3055
Sequence: 1 MCTTGLESLSLGRGAPTV.....HKLKYNHDLWENHEKLSLS:580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

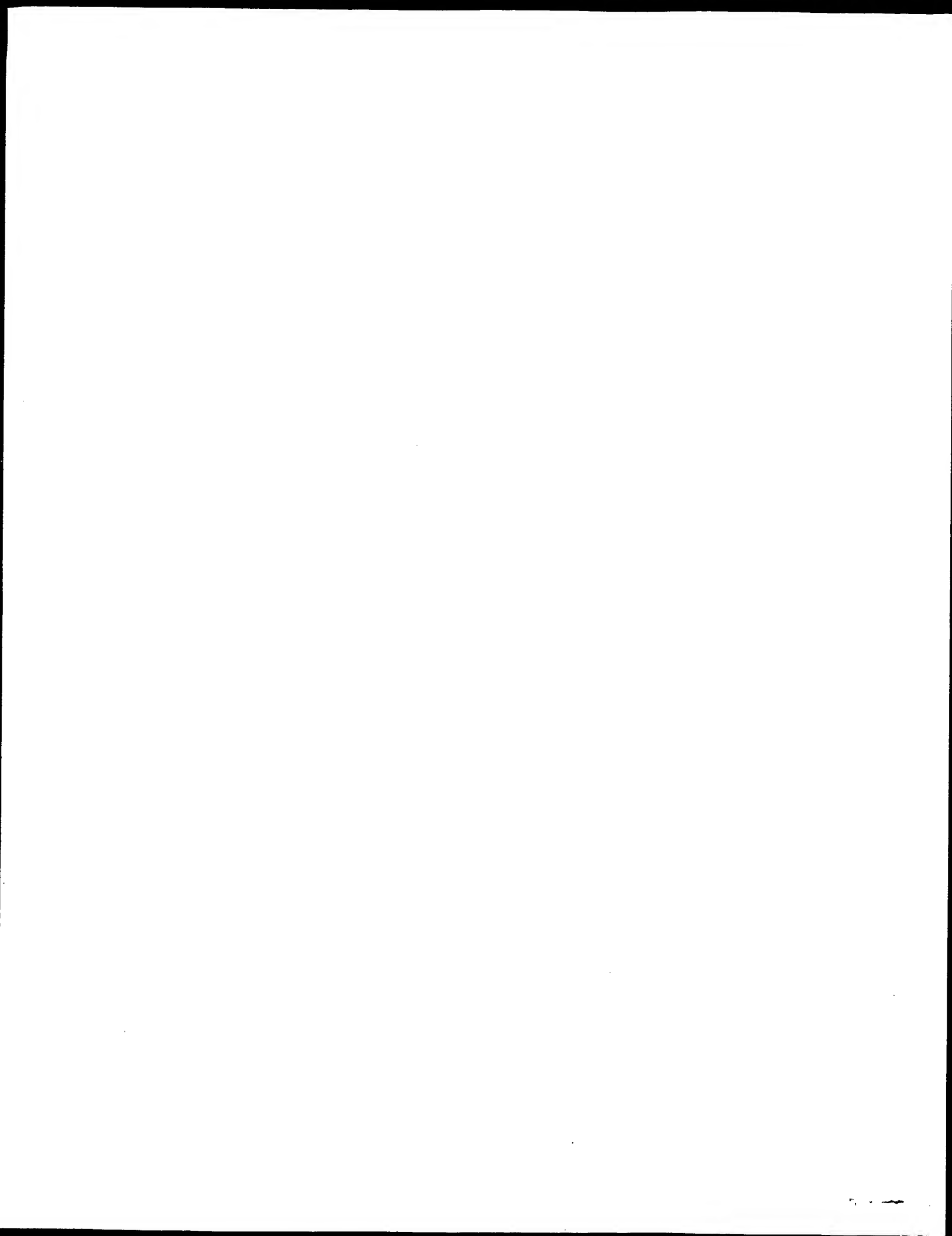
Result No.	Score	Query Match	Length	ID	Description
1	3055	100.0	580	AAW74473	Human MED1 endonuc
2	2958.5	96.8	565	AA144504	Human delta228-UV
3	2194.5	71.8	439	AA176548	Human ovarian tumo
4	1230	40.3	257	AAW88701	Secreted protein e
5	1194	6.4	219	AA120251	Human secreted pro
6	148.5	4.9	326	AA168859	Arabidopsis thalia
7	148.5	4.9	327	AA168859	Arabidopsis thalia
8	148.5	4.9	352	AA168857	Arabidopsis thalia
9	145	4.7	1644	AA125049	BRCA2 cancer susce
10	145	4.7	3418	AA192111	Human breast cance
11	145	4.7	3418	AAW23287	Human breast and o

12	145	4.7	3418	20	AAW04356	Human BRCA2 (omi3)
13	145	4.7	3418	20	AAW04357	Human BRCA2 (omi4)
14	145	4.7	3418	20	AAW04358	Human BRCA2 (omi5)
15	145	4.7	3418	20	AAW04354	Human BRCA2 (omi1)
16	145	4.7	3418	21	AAW77819	BRCA2 protein sequ
17	144	4.7	3418	18	AAW25038	Partial BRCA2 can
18	143	4.7	3418	20	AAW04355	Human BRCA2 (omi2)
19	141.5	4.6	782	11	AAW06991	Polypeptide antige
20	137	4.5	560	21	AAW19182	Lipid metabolism p
21	131.5	4.3	285	20	AAW14200	Mouse DNA demethyl
22	131.5	4.3	291	20	AAW14198	Human DNA demethyl
23	131	4.3	754	19	AAW33811	Tat stimulatory fa
24	130.5	4.3	411	20	AAW14197	Human DNA demethyl
25	130.5	4.3	411	22	AAW99915	Human protein sequ
26	130.5	4.3	411	22	AAW64314	Heart muscle cell
27	130.5	4.3	411	22	AAW64844	Expressed antigen
28	130	4.3	649	20	AAW90005	H. pylori GHPO 175
29	130	4.3	867	19	AAW98564	H. pylori GHPO 175
30	130	4.3	1743	19	AAW98879	Human prostate tum
31	127.5	4.2	281	20	AAW73829	Human prostate can
32	127.5	4.2	281	20	AAW48439	Human secreted pro
33	126	4.1	263	19	AAW74980	Mouse DNA demethyl
34	123	4.0	414	20	AAW14199	Peptide #4905 enco
35	122.5	4.0	995	22	AAW06223	Peptide #4097 enco
36	122.5	4.0	1202	22	AAW17663	Arabidopsis thalia
37	121.5	4.0	1024	21	AAW42350	Arabidopsis thalia
38	121.5	4.0	1105	21	AAW42349	Arabidopsis thalia
39	121.5	4.0	1189	21	AAW42348	Canine factor VIII
40	120.5	3.9	2343	20	AAW80989	Peptide #967 enco
41	120.5	3.9	2665	22	AAW14533	Peptide #987 enco
42	120.5	3.9	2665	22	AAW25950	Peptide #941 enco
43	120.5	3.9	2665	22	AAW02259	Human OREF ORF2255
44	120.5	3.9	3266	21	AAW42491	Plasmodium falcipa
45	119	3.9	646	21	AAW18188	

ALIGNMENTS

RESULT 1
AAW74473
ID AAW74473 standard; Protein; 580 AA.
XX AAW74473;
AC AAW74473;
XX AAW74473;
DT 19-MAY-1999 (first entry)
XX Human MED1 endonuclease protein sequence.
DE DE
DE DE
KW Endonuclease; MED1; human; methyl-CpG binding endonuclease-1;
KW DNA fidelity; DNA manipulation; cancer; fragile X syndrome; therapy;
KW myotonic dystrophy; Huntington's disease; spinocerebellar ataxia;
KW Kennedy's disease; triplet repeat expansion disorder.
XX Homo sapiens.
XX WQ9904636-A1.
PN WQ9904636-A1.
XX WQ9904636-A1.
PD 04-FEB-1999.
XX 04-FEB-1999.
PF 28-JUL-1998; 98WO-US15828.
XX 28-JUL-1998; 98WO-US15828.
XX 28-JUL-1997; 97US-0053936.
XX (FOX-) FOX CHASE CANCER CENT.
XX PI Bellacosa A;
XX WPI; 1999-142462/12.
DR N-PSDB; AAX22002#.
XX New nucleic acid encoding human endonuclease MED1 involved in DNA mismatch repair - used for diagnosing susceptibility to cancer and

Aug Nov 28/2000



fragile X syndrome, and therapeutically

Claim 8; Fig 3; 109pp; English.

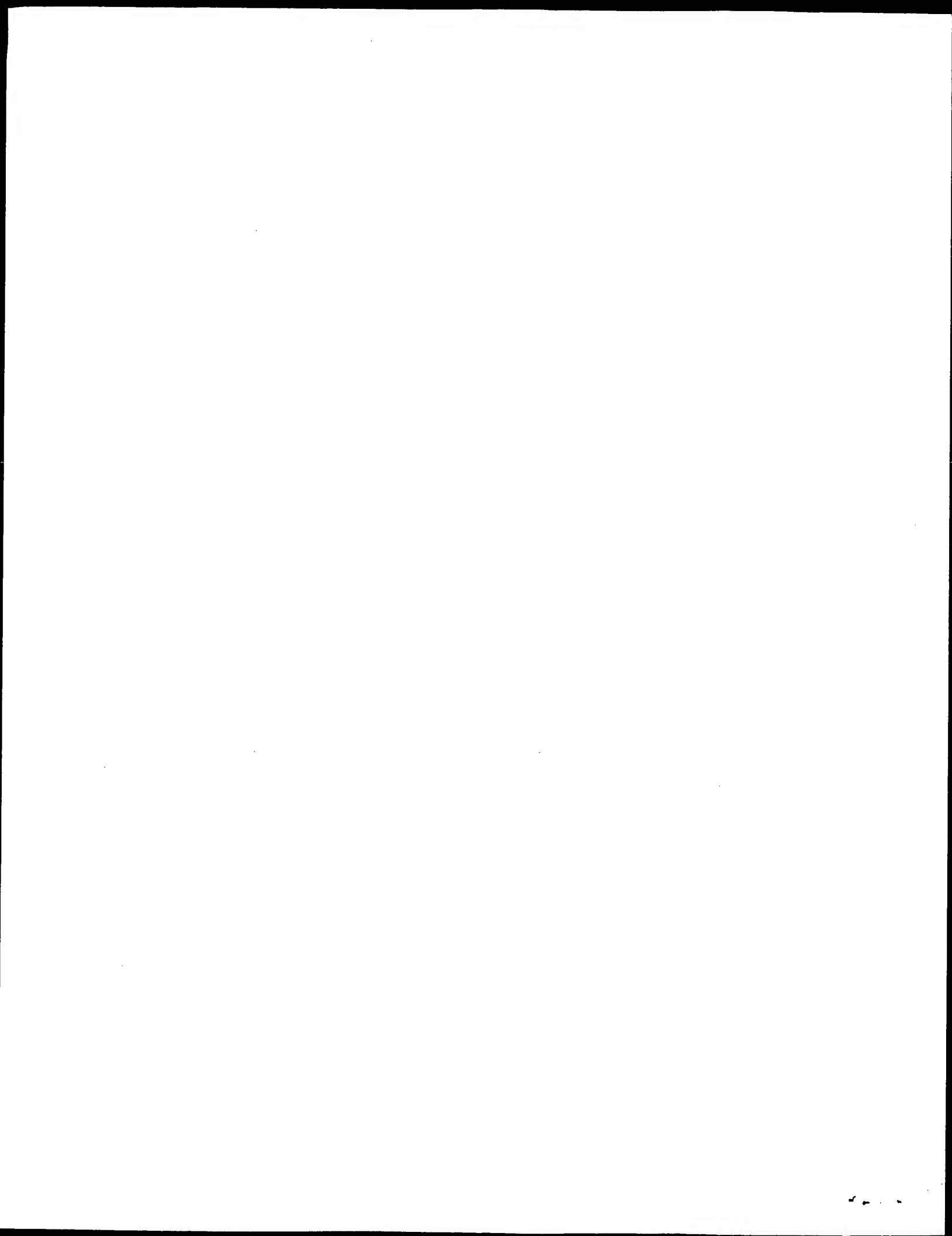
This sequence is the human MED1 endonuclease of the invention. MED1 (for methyl-CpG binding endonuclease-1) is used to screen for specific modulators (potential therapeutic agents particularly mimetics of MED1) and to study interactions involved in maintaining DNA fidelity, for DNA manipulation and to raise antibodies. Susceptibility or predisposition to cancer (particularly colorectal or endometrial, especially hereditary non-polyposis colorectal cancer), or its prognosis, where caused by alterations in the MED1-encoding gene, are identified by sequence comparison, amplification, detecting altered polypeptide, and restriction fragment mapping, hybridisation (particularly to probes specific for a mutant allele). These same methods can also be used to diagnose fragile X syndrome and other diseases (e.g. myotonic dystrophy, Huntington's disease, spinocerebellar ataxia and Kennedy's disease) associated with triplet repeat expansion. The DNA, or its fragments, are used as probes and primers in the above diagnostic methods, also to isolate homologous sequences, as sources of antisense sequences and for gene transfer, particularly to restore drug sensitivity to drug-resistant cancer cells.

Sequence 580 AA:

Query Match 100.0%; Score 3055; DB 20; Length 580;
Best Local Similarity 100.0%; Pred. No. 7.6e-284;
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGTGLESLSLGRGAAPTTSERLVPDPNDLRKEDVAMELERVGEDEQMIRKRS 120
DB 1 mgttgleslsldrgaaptvtsserlvppndlrkedvamelervgdeeqmmikrs 60
QY 61 CNPLLQEPPIASAFGATAGTECKRSVPCGWVRVKORLFGKTAGRFDVYFISPOGLKRS 120
DB 61 cnpllqepiasagfatagteckrsvpcgwerlvkqrlfgktagrfdvyfispqglkfrs 120
QY 121 KSSLANLHKNGETSLKPEDFDTVLKRGIKSRKDCSMAALTSHLQNSNNWNLRT 180
DB 121 ksslanlhkngetsllkpedfdftvlksrgiksrkdcsmaltshlqnqsnwnlrlt 180
QY 181 RSKCKDVFMPSSSSSELSQESRGLSNFTSHLLKDEGVDVDFNFRKVRKPKGVITLKG 240
DB 181 rskckdvfmppsssselsqesrglsnftshllkdegvddvfnfrkvrkpkgvitlkg 240
QY 241 IPKTKKGRKSCSGFVQSDSKRESVCNKADAEPVAKQSOLDRTVCISDAGACGETL 300
DB 241 ipkttkgrkscsgfvqsdskresvcnkadaesepvacksgldrtvcisdagacgetl 300
QY 301 SVTSEENSLVKKERSLSGSSNFCSEQKTSGLINKFCSAKDSHNEKYEDTFLESEEIGT 360
DB 301 svtseenslvkkerslsghssnfcseqktsglinkfcsakdshehnekyedtfleseeigt 360
QY 361 KVEVERKEHLHTDILKRSEMDNNSCPTKDFGTGKIFQEDTIPRTQIERKTSLYFSS 420
DB 361 kveverkehldilkrsemdnncsptrkdfgtgkifqedtiprtqierktslyfss 420
QY 421 KYNKEALSPRRKAFKWTTPSPFNLVQETLPHDPWKLIIATIFLNRTSKGMAIPVLWK 480
DB 421 kynkealspprrkafkwtppspfnlvqetlphdpwkliaatiflnrtskgmaipvlwk 480
QY 481 FLEKYPFAEARTADNRDYSLELKLPLGLYDLRAKTVKFTSDEVLTKOWKYPTELHGIGY 540
DB 481 flekypfaeartadnrdsellklplgllydlraktvkvftsdevltkowkyptelhgigy 540
QY 541 GNDYSIFVCVNEKQVHPEDHLKYNKYHDWLWENHEKLSLS 580
DB 541 gndysifvcvnekwqhpedhlnkynkyhdwlwenheklsls 580

ID AAY44504 standard; Protein; 565 AA.
XX AC AAY44504;
XX DT 27-MAR-2000 (first entry)
XX DE Human delta228-UV damage endonuclease.
XX KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide; glutathione-S-transferase signal peptide; uvell+ gene product; UV irradiation; DNA damage; UV radiation damage; photoproduct; abasic site; platinum diadduct; mismatched nucleotide pairing; nucleotide alkylation; skin cancer.
XX OS Homo sapiens.
XX PN WO9963828-A1.
XX PD 16-DEC-1999.
XX PF 08-JUN-1999; 99WO-US12910.
XX PR 08-JUN-1998; 98US-0088521.
XX PR 18-MAY-1999; 99US-0134752.
XX PA (UYEM-) UNIV EMORY.
XX PI Doetsch PW, Kaur B, Avery AM;
XX WPI: 2000-116417/10.
XX A new truncated ultraviolet damage endonuclease for treatment of skin cancers -
XX Claim 16; Page 60; 133pp; English.
XX The present sequence is human delta228-UV damage endonuclease.
XX Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the uvell+ gene product. This is expressed in frame with a GST leader sequence to generate a fusion protein. This provides stable endonuclease fragments for cleaving a double-stranded DNA molecule that has a distorted structure resulting from UV radiation damage, a photoproduct, an abasic site, mismatched nucleotide pairing, platinum diadduct, an intercalated molecule or alkylation of a nucleotide. Uvell can be used in compositions for internal or topical application and as a therapeutic agent for skin cancers.
XX SQ Sequence 565 AA;
Query Match 96.8%; Score 2958.5; DB 21; Length 565;
Best Local Similarity 97.4%; Pred. No. 1.3e-274;
Matches 565; Conservative 0; Mismatches 0; Indels 15; Gaps 1;
QY 1 MGTGLESLSLGRGAAPTTSERLVPDPNDLRKEDVAMELERVGEDEQMIRKRS 60
DB 1 mgttgleslsldrgaaptvtsserlvppndlrkedvamelervgdeeqmmikrs 60
QY 61 CNPLLQEPPIASAFGATAGTECKRSVPCGWVRVKORLFGKTAGRFDVYFISPOGLKRS 120
DB 61 cnpllqepiasagfatagteckrsvpcgwerlvkqrlfgktagrfdvyfispqglkfrs 120
QY 121 KSSLANLHKNGETSLKPEDFDTVLKRGIKSRKDCSMAALTSHLQNSNNWNLRT 180
DB 121 ksslanlhkngetsllkpedfdftvlksrgiksrkdcsmaltshlqnqsnwnlrlt 180
QY 181 RSKCKDVFMPSSSSSELSQESRGLSNFTSHLLKDEGVDVDFNFRKVRKPKGVITLKG 240
DB 181 rskckdvfmppsssselsqesrglsnftshllkdegvddvfnfrkvrkpkgvitlkg 240
QY 241 IPKTKKGRKSCSGFVQSDSKRESVCNKADAEPVAKQSOLDRTVCISDAGACGETL 300
DB 241 ipkttkgrkscsgfvqsdskresvcnkadaesepvacksgldrtvcisdagacgetl 300



Best Local Similarity 100.0%; Pred. No. 7.9e-190;
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRLKRNQISKRIVFTILKQAFKGNHPCVSVCTITYSRFHCLPDTLKSLPMSKT 60
DB 1 MLRLKRNQISKRIVFTILKQAFKGNHPCVSVCTITYSRFHCLPDTLKSLPMSKT 60

QY 61 TSLMLPQVNTGANSFSAETPVDLKKENETELANISGPHKKSTSTSRKRARSSKKKATDS 120
DB 61 TSLMLPQVNTGANSFSAETPVDLKKENETELANISGPHKKSTSTSRKRARSSKKKATDS 120

QY 121 VSKIDSVASYDSSTHLRSSRSKKPVNNSSESESEEQISKATKYKQKEEEVEE 180
DB 121 VSKIDSVASYDSSTHLRSSRSKKPVNNSSESESEEQISKATKYKQKEEEVEE 180

QY 181 VDESKLNESSDDEFVVPQLETPISKRRSSAKNEKESTMNLDHAPREMFDC 240
DB 181 VDESKLNESSDDEFVVPQLETPISKRRSSAKNEKESTMNLDHAPREMFDC 240

QY 241 DKPIPWGRGLGYACLTILRSKMERVFCSTRTITTTORGLSVKQLGTONVLDLILV 300
DB 241 DKPIPWGRGLGYACLTILRSKMERVFCSTRTITTTORGLSVKQLGTONVLDLILV 300

QY 301 EWNHNFGLHFRVSSDLFPFASHAKYGYTLFPAOSHLEEVGKLANYNHRLTWHPGGYTQ 360
DB 301 EWNHNFGLHFRVSSDLFPFASHAKYGYTLFPAOSHLEEVGKLANYNHRLTWHPGGYTQ 360

QY 361 IASPREVVVDSAIRDLAYHDEILSRMKLNEQNKDAVLIHLGTFEGKKTLDLRFK 420
DB 361 IASPREVVVDSAIRDLAYHDEILSRMKLNEQNKDAVLIHLGTFEGKKTLDLRFK 420

QY 421 QRLSDSVKARLVLENDVSVQDILLPLCQELNPLVLDWHHHNIVPGLTREGSLDMLPL 480
DB 421 QRLSDSVKARLVLENDVSVQDILLPLCQELNPLVLDWHHHNIVPGLTREGSLDMLPL 480

QY 481 IPTIRETWRTGKITOKOHYSADPTAISGMKRAHSDRVDFPDCDPTMDLMEAKEKE 540
DB 481 IPTIRETWRTGKITOKOHYSADPTAISGMKRAHSDRVDFPDCDPTMDLMEAKEKE 540

QY 541 QAVFELCRRYELQNPCCPLEIMGPEYDTRDGYPPGAEKRLTARKRRSRKEEVEDEK 599
DB 541 QAVFELCRRYELQNPCCPLEIMGPEYDTRDGYPPGAEKRLTARKRRSRKEEVEDEK 599

RESULT 2
Q01408 PRELIMINARY; PRT; 656 AA.

AC Q01408;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE OV-ENDONUCLEASE.
GN UVEL.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=95292980; PubMed=7774597;
RA Yajima H., Takao M., Yasuhira S., Zhao J.H., Ishii C., Inoue H.,
RA Yasui A.;
RT "A eukaryotic gene encoding an endonuclease that specifically repairs
DNA damaged by ultraviolet light."
RL EMBO J. 14:2393-2399(1995).
DR EMBL; D11392; CAB20113.1;
KW Endonuclease.
SQ SEQUENCE 656 AA; 74413 MW; 4B231E7A8735C86F CRC64;

Matches 241; Conservative 88; Mismatches 149; Indels 181; Gaps 18;

QY 110 ARSSKKKATDSVSKIDSVASYDSST-----HLRRSR-----SKPVNNSSESE- 157
DB 3 SRKSKAAALDTPQSESTFSSTLDSSAPARNLRSGRNILQSPSSKDRDHEKRSGEEL 62

QY 158 -----SEEQISKATKKVKQKEEEVEEVEDEK 184
DB 63 AGRMMGKQANGHCLREGKEQEGVKMATEGLARMERLQKQKOLEED----- 114

QY 185 SLKNSSDDEFVVPQLETPISKRRS-----RSSAKNEKESTMNLDH 232
DB 115 GIPVSVSRE-PTAPYHHKSTNAEREAKPEVLKTHSKDVEREAEIGDDVWMEPAAT 173

QY 233 ----PREMFDCLDK-----PIPWGRGLGYACLTILRSKMERVFCSTRT 275
DB 174 NIEPEDAQAAGAERPAVNSSYLPKWRGLGYACLTILRSKMERVFCSTRT 233

QY 276 TI-----ORDGLSVKQLGTONVLDLILKLVENHNF 307
DB 234 SIYDHRHLPQFEDEPEHHLKKNPKDSKEPQDELGHKFFQELGLANARDIVKMLCWNK 293

QY 308 IHFMRVSSDLFPFASHAKYGYTL-EFAOSHLEEVGKLANYNHRLTWHPGGYTQ 366
DB 294 IREFRLSSEMFPPASHPVGHYKGLAPFASEVLAEGRAVAELGHRLTTHPGQFTQLGSPRK 353

QY 367 VVDSAIRDLAYHDEILSRMKLNEQNKDAVLIHLGTFEGKKTLDLRFK 426
DB 354 EVESAIRDLAYHDEILSRMKLNEQNKDAVLIHLGTFEGKKTLDLRFK 413

QY 427 VKARLVLENDVSVQDILLPLCQELNPLVLDWHHHNIV--PGTLREGSLDML-PLIP 482
DB 414 CKNLVLENDVSVQDILLPLCQELNPLVLDWHHHNIV--PGTLREGSLDML-PLIP 473

QY 483 TIRETWRTGKITOKOHYSADPTAISGMKRAHSDRVDFPDCDPTMDLMEAKEKE 542
DB 474 RIANTWKRGKITOKOHYSADPTAISGMKRAHSDRVDFPDCDPTMDLMEAKEKE 532

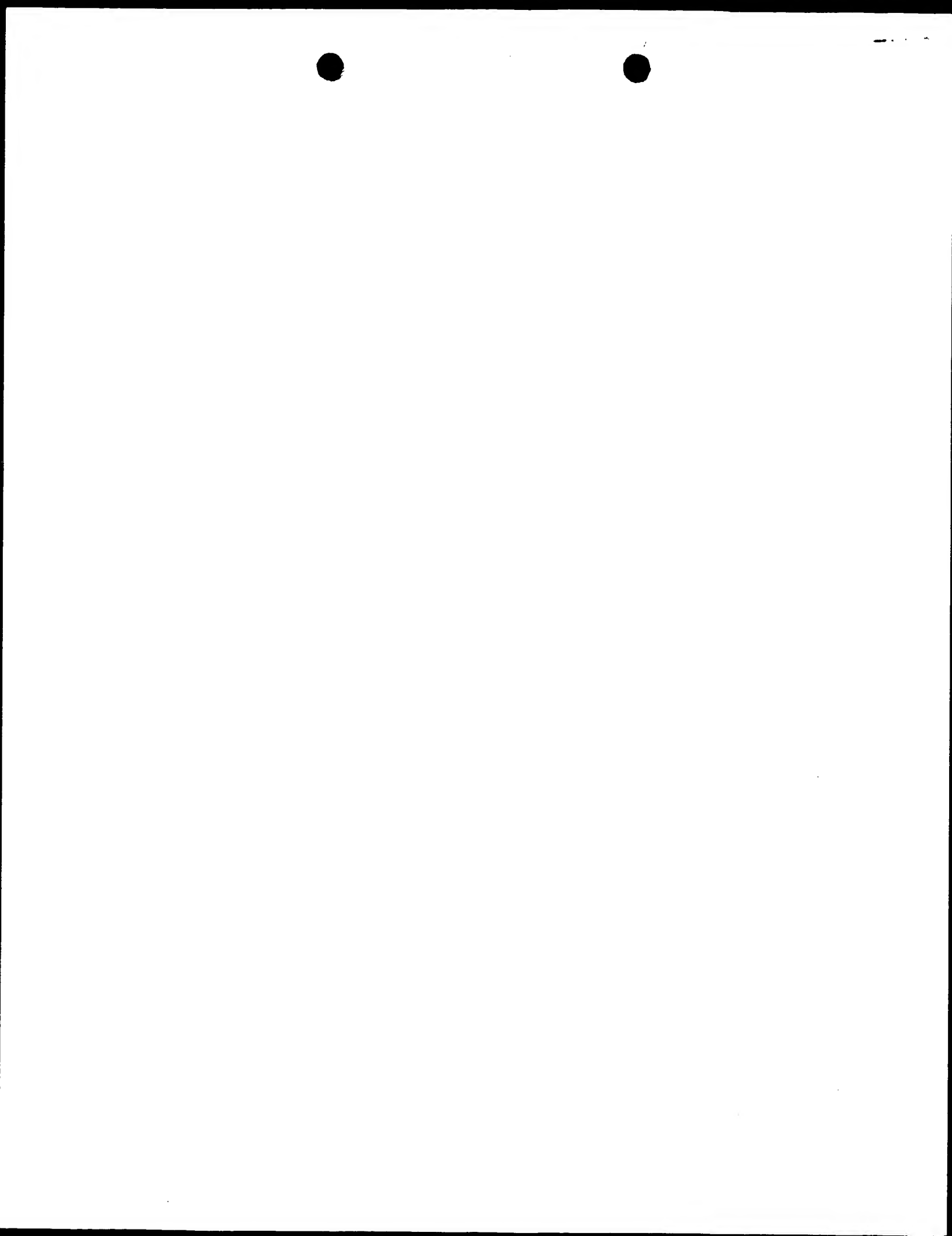
QY 543 VFELCRRYEL-----ONPPCP-----LEIMGPE--- 565
DB 533 VFELCRRYELQNPCCPLEIMGPEYDTRDGYPPGAEKRLTARKRRSRKEEVEE 592

QY 566 --YDQTRDG-----YYPGGAEKRLTARKRRSRK-----EEVEE 597
DB 593 TAADDVKDAPEGKPEVEBERAMGPGYNNVWPLGCEEWLKPCKREVKKGVPEVEDE 651

RESULT 3
Q014F0 PRELIMINARY; PRT; 317 AA.

AC Q014F0;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PUTATIVE UV-ENDONUCLEASE.
GN YWJD.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14579 TYPE STRAIN;
RX MEDLINE=200555637; PubMed=10589720;
RA Okstad O., Gominet M., Purnelle B., Rose M., Lereclus D., Kolsto A.B.;
RT "Sequence analysis of three Bacillus cereus loci under P1cR-regulated
genes encoding degradative enzymes and enterotoxin."
RL Microbiology 145:3129-3138(1999).
DR EMBL; AJ243712; CAB69813.1;
KW Endonuclease.
SQ SEQUENCE 317 AA; 36990 MW; A6AC35F5800A22C3 CRC64;

Query Match 32.5%; Score 1011.5; DB 3; Length 656;
Best Local Similarity 36.6%; Pred. No. 2.2e-56;



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:52:23 ; Search time 74.94 Seconds
(without alignments)
666.807 Million cell updates/sec

Title: US-09-724-296-36
Perfect score: 3482
Sequence: 1 MPRSKSAAALDTPOSESST.....REVKKGKVPVEVEGEFGD 656

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

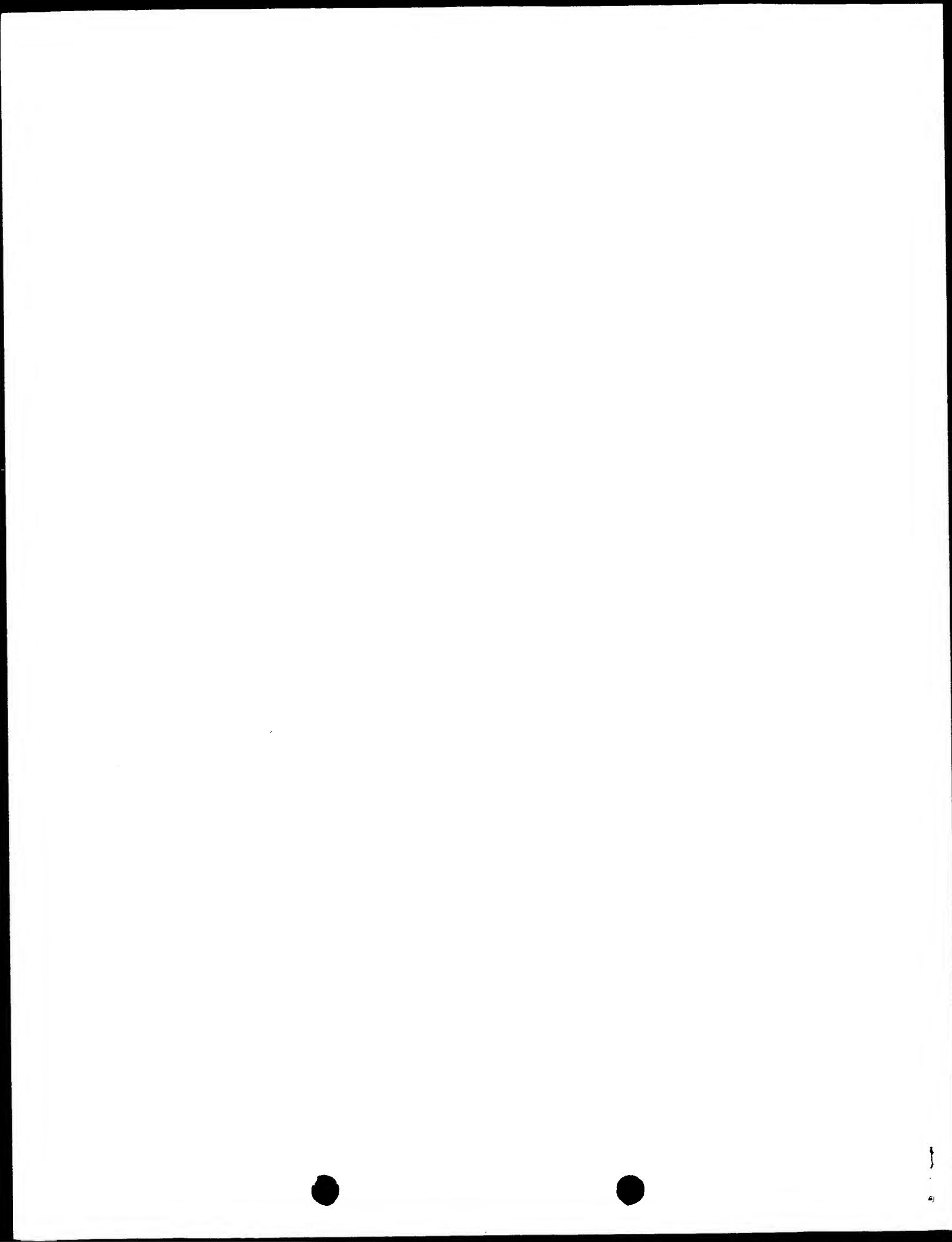
Result No.	Score	Query Match	Length	ID	Description
1	3472	99.7	656	S55262	UV-endonuclease -
2	1011.5	29.0	599	S71134	UV-endonuclease -
3	377	10.8	322	E83974	hypothetical prote
4	348.5	10.0	320	S55418	UV-endonuclease ho
5	261.5	7.5	326	C75350	probable UV damage
6	160.5	4.6	1280	T00365	hypothetical prote
7	155.5	4.5	5327	T13564	microtubule-associ
8	146	4.2	796	S56231	hypothetical prote
9	141.5	4.1	1110	I51116	NF-180 - sea lamp
10	140	4.0	727	T17292	hypothetical prote
11	139.5	4.0	1510	T33100	hypothetical prote
12	139	4.0	1815	C81169	IgA-specific metal
13	138	4.0	1547	T28657	blackjack protein, p
14	137	3.9	1560	T42727	cell proliferation
15	137	3.9	2897	B48666	cell proliferation
16	137	3.9	3256	A48666	cell proliferation
17	136.5	3.9	845	A45669	neurofilament trip
18	136.5	3.9	849	S00030	neurofilament trip
19	136	3.9	3942	T42730	Bassoon protein -
20	135.5	3.9	791	T24866	hypothetical prote
21	135	3.9	793	JH0628	caldesmon - human
22	134	3.8	2663	S28261	centromere protein
23	133.5	3.8	980	E71606	hypothetical prote
24	132.5	3.8	915	S54485	CEB1 protein - yea
25	132.5	3.8	6642	T29375	hypothetical prote
26	132	3.8	734	F85073	protein UNC-89 - C
27	131.5	3.8	1222	T22490	hypothetical prote
28	131	3.8	606	A43427	neurofilament trip
29	128.5	3.7	856	T16543	hypothetical prote

30 128.5 3.7 7962 2 I38346 elastic titin - hu
31 128 3.7 1020 1 QFH0H neurofilament trip
32 128 3.7 1805 2 A34736 nestin - rat
33 128 3.7 3488 2 T34418 hypothetical prote
34 128 3.7 4151 2 T13734 groovin gene prote
35 127.5 3.7 2453 2 S60254 nuclear receptor c
36 126.5 3.6 720 2 T36819 hypothetical prote
37 126.5 3.6 2416 2 T13825 adenomatous polypo
38 126 3.6 522 2 C96608 hypothetical prote
39 126 3.6 1773 2 A81937 IgA-specific metal
40 125.5 3.6 786 2 C86406 hypothetical prote
41 125.5 3.6 1400 2 T52359 hypothetical prote
42 125.5 3.6 2441 2 D71623 erythrocyte membra
43 124.5 3.6 1829 2 T26135 hypothetical prote
44 124.5 3.6 4377 2 A55575 ankyrin 3, long sp
45 124 3.6 1072 1 A37221 neurofilament trip

ALIGNMENTS

RESULT 1
S55262
UV-endonuclease - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
C:Accession: S55262
R:Yajima, H.; Takao, M.; Yasuhira, S.; Zhao, J.H.; Ishii, C.; Inoue, H.; Yasui, A.
EMBO J. 14, 2393-2399, 1995
A:Title: A eukaryotic gene encoding an endonuclease that specifically repairs DNA dam
A:Reference number: S55262; MUID:95292980
A:Accession: S55262
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-656 <YAU>
A:Cross-references: EMBL:D11392; NID:g1526560; PID:g927215
C:Genetics:
A:Gene: uvel

Query Match 99.7%; Score 3472; DB 2; Length 656;
Best Local Similarity 99.7%; Pred. No. 2.2e-208;
Matches 654; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MPRSKSAAALDTPOSESSTFSSLDSSAPSPARNLRSGRNILQPSSEKDRDHEKRSGE 60
Db 1 MPRSKSAAALDTPOSESSTFSSLDSSAPSPARNLRSGRNILQPSSEKDRDHEKRSGE 60
Qy 61 ELAGRMGKDANGHCLREGKEQEGVKMAIEGLARMERRLQQRATKROKKOLEEDGIPVPS 120
Db 61 ELAGRMGKDANGHCLREGKEQEGVKMAIEGLARMERRLQQRATKROKKOLEEDGIPVPS 120
Qy 121 VVSREPTAPYHHKSTNAFEREAKPEVLKTHSKOYERAEIGVDDVVKMEPAATNIIEPED 180
Db 121 VVSREPTAPYHHKSTNAFEREAKPEVLKTHSKOYERAEIGVDDVVKMEPAATNIIEPED 180
Qy 181 AODAEAGAAPPAVNSYLPMPKGRGLVACLTNTYLRNAKPPFSSRTCTMASIVDHRH 240
Db 181 AODAEAGAAPPAVNSYLPMPKGRGLVACLTNTYLRNAKPPFSSRTCTMASIVDHRH 240
Qy 241 PLOFEDEPEHHLKNDKSKPEQDELGHKFVQELGLANARDIVKMLCWNKYGIRFLRLS 300
Db 241 PLOFEDEPEHHLKNDKSKPEQDELGHKFVQELGLANARDIVKMLCWNKYGIRFLRLS 300
Qy 301 SEMPPFASHPVHGYKLAPFASEVLAEAGRAVAELGHRLLTHPGQFTQLGSPRKEVESA 360
Db 301 SEMPPFASHPVHGYKLAPFASEVLAEAGRAVAELGHRLLTHPGQFTQLGSPRKEVESA 360
Qy 361 RDLFYHDELLSLKLPQQNRDAVMIIHMGCGFGDKAATLERFKRNYARLSQSKNRLVL 420
Db 361 RDLFYHDELLSLKLPQQNRDAVMIIHMGCGFGDKAATLERFKRNYARLSQSKNRLVL 420
Qy 421 ENDDVGWTVHDLPLPVCEELNIPMVLDYHHHNCIFDPAHLREGTLDISDPKLOERIANWK 480



Db 421 ENDVGTVDHLLPVCEELNIPVLDYHHNICFPAHLREGTLDISDPKLOERIANTWK 480
 QY 481 RGIKOKMYSEPCDAGVTPDRRRKRRPVMTPPCPPDMDLMIEAKKEQAVFELMRTF 540
 Db 481 RGIKOKMYSEPCDAGVTPDRRRKRRPVMTPPCPPDMDLMIEAKKEQAVFELMRTF 540
 QY 541 KLPGEKINDMYPYRDDNRRPAPVPKPKKGGKRRKRTTDEAAAEPEVDTAADDVKD 600
 Db 541 KLPGEKINDMYPYRDDNRRPAPVPKPKKGGKRRKRTTDEAAAEPEVDTAADDVKD 600
 QY 601 APEGKPEVEERAMGGPNRYVWPLGCEEWLKPCKREVKKGVPEEVEDGEFGD 656
 Db 601 APEGKPEVEERAMGGPNRYVWPLGCEEWLKPCKREVKKGVPEEVEDGEFGD 656
 RESULT 2
 S71134
 UV-endonuclease - fission yeast (Schizosaccharomyces pombe)
 Species: Schizosaccharomyces pombe
 Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 20-Jun-2000
 Accession: S71134; T39815
 R.Takao, M.; Yonemasu, K.; Yamamoto, K.; Yasui, A.
 Nucleic Acids Res. 24, 1267-1271, 1996
 A:Title: Characterization of a UV endonuclease gene from the fission yeast Schizosaccharomyces pombe
 A:Reference number: S71134; MUID:96188860
 A:Accession: S71134
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-599 <TAK>
 A:Cross-references: EMBL:D78571; NID:g1399000; PIDN:BAAL1415.1; PID:g1399001
 A:Experimental source: strain SP972
 R.Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Lucas, M.; Gaillardin, C.
 submitted to the EMBL Data Library, June 1998
 A:Reference number: 221881
 A:Accession: T39815
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-599 <LYN>
 A:Cross-references: EMBL:AL023859; PIDN:CAAL19577.1; GSPDB:GN00067; SPDB:SPBC19C7.09c
 A:Experimental source: strain 972h; cosmid c19C7
 C:Genetics:
 A:Gene: UVDE; SPBC19C7.09c
 A:Map position: 2

Query Match 29.0%; Score 1011.5; DB 2; Length 599;
 Best Local Similarity 36.6%; Pred. No. 1.9e-55;
 Matches 241; Conservative 88; Mismatches 149; Indels 181; Gaps 18;
 QY 3 SRKSAALDTPSESTFSTLSSAPSPARNLRSGRNILQPSKEDRDHEKRSSEEL 62
 Db 110 ARSKKKATDSVDKIDESVASYDSST-----HLRSSR-----SKKPNYNSSESE- 157
 QY 63 AGRMGRDANGCHLRECKEGBEGVMAIEGLARMERLRORATKROKQLEED----- 114
 Db 158 -----SEQISKATKKVKQKEEYVEEVDK 184
 QY 115 GIPVPSVYSRF-PTAPYHHKSTNAEREAKEPEVLTKTHSKDVEEAEIGVDVVVKMEPAT 173
 Db 185 SLKNSSSDEEPVVPQLETPISKRRS-----RSANLEKESMTNLDHHA----- 232
 QY 174 NIEPEDAQDAERGAARPPAVNNSYLPLPWKGLGYACILTYLRNAKPPIFSSRTCRMA 233
 Db 233 -----PREMFCDLCK-----PIPWGRGLGYACILTYLRNAKPPIFSSRTCRIT 275
 QY 234 SIYDHRHPLQFEDEPEHHLKPKDSKEQDELGHKFEVQELGLANARIVKMLCWNRYG 293
 Db 276 TI-----QRDGLSVRLQGTQNVLDLILKLVENHNFG 307
 QY 294 IRELRLSSEFPFASHVHYGKILAPPASEVLAEGRAVAAELGHLRTHHPGQFTQLGSPRK 353
 Db 308 IHPMYRSSDLFPFASHAKYGYTL-EFAQSHLEVGLKANKYNNHRLTMHPGQYTIASPRE 366

QY 354 EVVESAIRDLEYHDELLSLKLPEQQNRDAMIHHMGQFGDKAATLERFKRYARLSQS 413
 Db 367 VVDSAIRDLAYHDEILSRMLKNEQLNKDAVLIHLCGTPEGKKTDLDRFKNYQLRSDS 426
 QY 414 CKNRLVLENDVGTVDHLLPVCEELNIPVLDYHHNICFPAHLREGTLDISDPKLOE 473
 Db 427 VKARVLENDVSVQDILLPLCQELNIPVLWVHHNIV--PGTREGSLDLM--PLIP 482
 QY 474 RIANTWKRGIKOKMYSEPCD-GAVTPDRRRKRRPVMTPPCPPDMDLMIEAKKEQA 532
 Db 483 TIRETWRTKGIKOKMYSEADPTAIGMKRRAHSRVDTFPCDPTMDLMIEAKKEQA 542
 QY 533 VFELMRTFKLPGEKINDMYPYRDDNRRPAPVPKPKKGGKRRKRTTDEAAAEPEVD 592
 Db 543 VFELCRRYEL-----ONPPCP-----LEIMGPE--- 565
 QY 593 TAADDVKDAPGKPEVEERAMGGPNRYVWPLGCEEWLKPCKREVKKGVPEEVEDE 651
 Db 566 --YDQTRDG-----YYPGAEKRLTARKRSRK-----EVEED 597
 RESULT 3
 E83974
 hypothetical protein BH2597 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
 C:Accession: E83974
 R.Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Fujii, F.; H.
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
 submitted to the EMBL Data Library, June 1998
 A:Reference number: A83650; MUID:20263314
 A:Accession: E83974
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-322 <STO>
 A:Cross-references: GB:AF001516; GB:BA000004; NID:g10175192; PIDN:BA06316.1; GSPDB:G
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH2597

Query Match 10.8%; Score 377; DB 2; Length 322;
 Best Local Similarity 29.1%; Pred. No. 2.4e-16;
 Matches 105; Conservative 61; Mismatches 127; Indels 68; Gaps 10;
 QY 207 RUGYACLTLYLRNAKPPIFSSRTCRMASIVDHRHPLQFEDEPEHHLKKNPKDSKEPQDEL 266
 Db 4 QGYVAMSWELANASP-----SKTMTAT-----QEKIEDH-----EA 36
 QY 267 GHKFVQELGLANARDIVKMLCWNKEYGIRFLRLSSEMPFASHPV-HGYKLAPFASEVLA 325
 Db 37 GLRKLERIAKTNLHCLRLKHLNLAYQISFRLSSKLVPLVNHPLTEGWKYLATAEELQ 96
 QY 326 EAGRVAELGLHRLTHPCQFTQLGSPKRVVESAIRDLEYHDELLSLKLPEQQNRDAMV 385
 Db 97 AVGEFASCHQRIDFHPDHFVVLNASEAKITRRSLQTLHYHKLKGMIDPRHR-----C 152
 QY 386 IHHMGQFGDKAATLERFKRYARLSOSCKNRLVLENDVGTVDHLLPVCEELNIPWL 445
 Db 153 VLHVGKKKGVEAGLEQFIENTASIPKLSLMIENLNDKSYTIDVLYLGERLAIPWL 212
 QY 446 DYHHNICFDPFAHLREGTLDISDPKLOERIANTWKRKGIKOKMYSEPCDGAVTPTPR 501
 Db 213 DIHHHDVHLHRSKSLQE-----TWQRIVATWEDSPLPVKIHLSPLSGEDDPRHHDY 263
 QY 502 --DRRKRHRPVMTPPCPPD-----MDLMEAKKEQAVFELMRTFKLPGEKINDMYPY 554
 Db 264 INADR-----FIAFLHETGADAVDHLHVMIEAKKDLALFQLMK-----DLAEY 307
 QY 555 D 555
 Db 308 D 308



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:06:28 ; Search time 134.53 seconds
(without alignments)
630.625 Million cell updates/sec

Title: US-09-724-296-38

Perfect score: 3055

Sequence: 1 MGTGLESLSIGRGAAPTV.....HKLKHYHDLWLENHEKLSLS 580

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mhc.*
7: sp_mammal.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_invertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3055	100.0	580	095243	095243 homo sapien
2	1821.5	59.6	554	11 Q9Z2D7	Q9Z2D7 mus musculus
3	880.5	28.8	416	13 Q919F1	Q919F1 gallus gall
4	344	11.3	419	10 Q9SFC1	Q9SFC1 arabidopsis
5	220.5	7.2	467	13 Q9YGC6	Q9YGC6 xenopus lae
6	202.5	6.6	344	13 Q42403	Q42403 gallus gall
7	146	4.8	702	10 Q9SF22	Q9SF22 arabidopsis
8	143.5	4.7	782	5 Q25875	Q25875 plasmodium
9	141.5	4.6	782	5 Q25730	Q25730 plasmodium
10	141.5	4.6	782	5 Q26007	Q26007 plasmodium
11	140.5	4.6	782	5 Q9U430	Q9U430 plasmodium
12	140.5	4.6	782	5 Q9U414	Q9U414 plasmodium
13	138.5	4.5	782	5 Q9U429	Q9U429 plasmodium
14	137.5	4.5	782	5 Q26104	Q26104 plasmodium
15	137.5	4.5	2209	5 Q9U0G6	Q9U0G6 plasmodium
16	136.5	4.5	782	5 Q9U431	Q9U431 plasmodium
17	136.5	4.5	1359	5 O02061	O02061 caenorhabdi
18	135.5	4.4	282	13 Q9PUM9	Q9PUM9 xenopus lae
19	135.5	4.4	755	4 O43719	O43719 homo sapien

20 135.5 4.4 1058 5 09V433
21 135 4.4 822 3 09USH9
22 133.5 4.4 1359 5 09U7E0
23 132.5 4.3 286 1 09YDP0
24 132 4.3 1819 2 09ZLV0
25 131.5 4.3 230 1 09P9L6
26 131.5 4.3 285 11 09Z2D8
27 131.5 4.3 291 4 095983
28 131 4.3 754 4 099730
29 130.5 4.3 411 4 09UBB5
30 130 4.3 537 5 09VZT4
31 130 4.3 1927 2 025262
32 129.5 4.2 3122 11 09JMD6
33 129 4.2 1163 4 09UBH7
34 128.5 4.2 946 10 09C9D8
35 128 4.2 853 10 09LFE4
36 127.5 4.2 1058 5 09U5E0
37 127 4.2 648 10 09LQV6
38 127 4.2 970 10 09AVF2
39 127 4.2 1723 2 09JMX8
40 126.5 4.1 1183 5 09W3D1
41 126 4.1 303 13 09PUM8
42 126 4.1 561 10 09C6S8
43 126 4.1 881 5 09BKN8
44 125.5 4.1 302 4 09UIS8
45 125 4.1 873 11 09WVC9

ALIGNMENTS

RESULT 1
095243
ID 095243 PRELIMINARY; PRT; 580 AA.
AC 095243;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE METHYL-CPG BINDING PROTEIN MBD4.
GN MBD4 OR MBD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449942; PubMed=9774669;
RA Hendrich B., Bird A.;
RT "Identification and characterization of a family of mammalian methyl-
CPG binding proteins.";
RA Mol. Cell. Biol. 18:6538-6547(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=FETAL BRAIN;
RX MEDLINE=99199294; PubMed=10097147;
RA Bellacosa A., Cicchilitti L., Schepis F., Riccio A., Yeung A.T.,
Matsumoto Y., Golenis E.A., Genuardi M., Neri G.;
RT "MED1, a novel human methyl-CpG-binding endonuclease, interacts with
DNA mismatch repair protein MLH1.";
EL Proc. Natl. Acad. Sci. U.S.A. 96:3969-3974(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99373255; PubMed=10441743;
RA Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;
RT "Genomic structure and chromosomal mapping of the murine and human
mbd1, mbd2, mbd3, and mbd4 genes.";
RL Mamm. Genome 10:906-912(1999).
DR EMBL; AF172250; AAC68879.1; -;
DR EMBL; AF114784; AAD22195.1; -;
DR EMBL; AF120999; AAD50374.1; -;
DR EMBL; AF120997; AAD50374.1; JOINED.
DR EMBL; AF120998; AAD50374.1; JOINED.
DR InterPro; IPR001739; MBD.



DR InterPro: IPR003265; Endo_3c.
DR Pfam: PF00730; HhH-GPD; 1.
DR Pfam: PF01429; MBD; 1.
DR SMART: SM00391; MBD; 1.
KW Endonuclease.
SQ SEQUENCE 580 AA; 66050 MW; BF16FB21A34B8E5F CRC64;

Query Match 100.0%; Score 3055; DB 4; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.9e-219;
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGTGLESLSLGRGAAPTTSERLVPPDPNDLRKEDVAMELERVGEDEEOMMKRSSE 60
DB 1 MGTGLESLSLGRGAAPTTSERLVPPDPNDLRKEDVAMELERVGEDEEOMMKRSSE 60
QY 61 CNPLLEPTASAFQATAGTECKRSVPCGWERVVKORLFGKTAGRFDVYFISPOGLKFRS 120
DB 61 CNPLLEPTASAFQATAGTECKRSVPCGWERVVKORLFGKTAGRFDVYFISPOGLKFRS 120
QY 121 KSLANYLHKNGETSLKPEDFDFTVLSKRGKISRYKDCSMAALTSHLQNSNNMNLRT 180
DB 121 KSLANYLHKNGETSLKPEDFDFTVLSKRGKISRYKDCSMAALTSHLQNSNNMNLRT 180
QY 181 RSKCKKDVMPSPSSSELOESRGLSNFTSTHLLKREDEGVDDVNFRRKPKGKVTILKG 240
DB 181 RSKCKKDVMPSPSSSELOESRGLSNFTSTHLLKREDEGVDDVNFRRKPKGKVTILKG 240
QY 241 IPIKTKKCGKSCGFSQSDSKRESVCNKADAESPVAQKSQLDRTVCISDAGACGETL 300
DB 241 IPIKTKKCGKSCGFSQSDSKRESVCNKADAESPVAQKSQLDRTVCISDAGACGETL 300
QY 301 SVTSEENSLVKKERSLSGSGNFCSEQKTSGIINKFCSAKDSHNEKYEDTFLSEEEIGT 360
DB 301 SVTSEENSLVKKERSLSGSGNFCSEQKTSGIINKFCSAKDSHNEKYEDTFLSEEEIGT 360
QY 361 KVEVERKEHLHTDILKRGSEMDNNSCPTKQFTGKIFQEDTIPRTQIERKRTSLYFSS 420
DB 361 KVEVERKEHLHTDILKRGSEMDNNSCPTKQFTGKIFQEDTIPRTQIERKRTSLYFSS 420
QY 421 KYNKEALSPRRKAKFKKWTTPRSPFNVLQETLFDHPWKLLIATIFLNRTSGKMAIPVLWK 480
DB 421 KYNKEALSPRRKAKFKKWTTPRSPFNVLQETLFDHPWKLLIATIFLNRTSGKMAIPVLWK 480
QY 481 FLEKPSAEVARTADRDVSELLKPLGLYDLRAKTIIVKFSDEYLTQWKYPIELHGIGKY 540
DB 481 FLEKPSAEVARTADRDVSELLKPLGLYDLRAKTIIVKFSDEYLTQWKYPIELHGIGKY 540
QY 541 GNDSYRIFCVNEWKQVHPEDHKLKYNKHDWLNWENHEKLSLS 580
DB 541 GNDSYRIFCVNEWKQVHPEDHKLKYNKHDWLNWENHEKLSLS 580

RESULT 2

Q92D7 PRELIMINARY; PRT; 554 AA.
ID Q92D7;
AC Q92D7;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE METHYL-CPG BINDING PROTEIN MBD4.
GN MBD4.
OS Mus musculus (Mouse), and
OS Mus musculus domesticus (western European house mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090, 10092;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449942; PubMed=9774669;
RA Hendrich B., Bird A.;
RT "Identification and characterization of a family of mammalian methyl-
CPG binding proteins.";

RL Mol. Cell. Biol. 18:6538-6547(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=99373255; PubMed=10441743;
RA Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;
RT "Genomic structure and chromosome mapping of the murine and human
Mamm. Genome 10:906-912(1999).
RL ENBL; AF120996; AAC68878.1; -;
DR ENBL; AF120996; AAC68878.1; -;
DR MGD; MGI:1333850; Mbd4.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR001739; MBD.
DR InterPro: IPR002052; N6_Mtase.
DR Pfam: PF00730; HhH-GPD; 1.
DR Pfam: PF01429; MBD; 1.
DR SMART: SM00391; MBD; 1.
DR PROSITE: PS00092; N6_MTASE; UNKNOWN_1.
SQ SEQUENCE 554 AA; 62577 MW; 792D37CB180291F5 CRC64;

Query Match 59.6%; Score 1821.5; DB 11; Length 554;
Best Local Similarity 66.2%; Pred. No. 1.4e-127;
Matches 384; Conservative 49; Mismatches 116; Indels 31; Gaps 11;

QY 6 LESLSLGD---RGAAPTVTSSERLVPPDPNDLRKEDVAMELERVGEDEEOMMKRSSECN 62
DB 1 MESPLGDNVRVG-----ESLVPPDPNDLRKEDVAMELERVGEDEEOMMKRSSECN 62
QY 63 PLLEPTASAFQATAGTECKRSVPCGWERVVKORLFGKTAGRFDVYFISPOGLKFRSKS 122
DB 51 SLLOEPTAST--LSSTATTATGHPVPCGWERVVKORLFGKTAGRFDVYFISPOGLKFRSKR 109
QY 123 SLANYLHKNGETSLKPEDFDFTVLSKRGKISRYKDCSMAALTSHLQNSNNMNLRTS 182
DB 110 SLANYLHKNGETSLKPEDFDFTVLSKRGKISRYKDCSMAALTSHLQNSNNMNLRTS 169
QY 183 KCKKDVMPSPSSSELOESRGLSNFTSTHLLKREDEGVDDVNFRRKPKGKVTILKGIP 242
DB 170 KWKTDVLPSPSGTSPSSSELOESRGLSNFTSTHLLKREDEGVDDVNFRRKPKGKVTILKGTA 229
QY 243 IKTKKCGKSCGFSQSDSKRESVCNKADAESPVAQKSQLDRTVCISDAGACGETLSV 302
DB 230 SOKTKQKCKSLLESTQNRNRKRSVQKVGADRELVPQESQLNFTLCPADACA--RETUGL 288
QY 303 TSEENSLVKKERSLSGSGNFCSEQKTSGIINKFCSAKDSHNEKYEDTFLSEEEIGTKV 362
DB 289 AGE-----EKSPFGLDLCFIQVTSGTNKPSTEAAGEANR--EQTFLESEIRSK - 338
QY 363 EVVERK--EHLHTDILKRGSEMDNNSCPTKQFTGKIFQEDTIPRTQIERKRTSLYFSS 420
DB 339 --GDRKGEAHLHTGVLDQSEMP--SCSQAKKHFTSE--TFQEDSIPRTQVEKRTSLYFSS 394
QY 421 KYNKEALSPRRKAKFKKWTTPRSPFNVLQETLFDHPWKLLIATIFLNRTSGKMAIPVLWK 480
DB 395 KYNKEALSPRRKAKFKKWTTPRSPFNVLQETLFDHPWKLLIATIFLNRTSGKMAIPVLWK 454
QY 481 FLEKPSAEVARTADRDVSELLKPLGLYDLRAKTIIVKFSDEYLTQWKYPIELHGIGKY 540
DB 455 FLEKPSAEVARTADRDVSELLKPLGLYDLRAKTIIVKFSDEYLTQWKYPIELHGIGKY 514
QY 541 GNDSYRIFCVNEWKQVHPEDHKLKYNKHDWLNWENHEKLSLS 580
DB 515 GNDSYRIFCVNEWKQVHPEDHKLKYNKHDWLNWENHEKLSLS 554
RESULT 3
Q919F1 PRELIMINARY; PRT; 416 AA.
ID Q919F1;
AC Q919F1;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:45:51 ; Search time 74.94 Seconds
(without alignments)
841.640 Million cell updates/sec

Title: US-09-724-296-2
Perfect score: 4349
Sequence: 1 MTKLPILGYWKIKGLVPTTR.....KRLTARKRRSRKEVEDEK 828
Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3112	71.6	599	2 S71134	UV-endonuclease -
2	1184	27.2	219	2 A26484	glutathione transf
3	1011.5	23.3	656	2 S55262	UV-endonuclease -
4	984	22.6	218	2 A45556	glutathione S-tran
5	921	21.2	218	2 A45523	glutathione transf
6	654	15.0	209	2 A48388	glutathione S-tran
7	481.5	11.1	218	2 S33860	glutathione transf
8	481.5	11.1	218	2 A29794	glutathione transf
9	478.5	11.0	218	2 S32425	glutathione transf
10	477.5	11.0	218	2 S65674	glutathione transf
11	474.5	10.9	218	2 A47486	glutathione transf
12	473.5	10.9	217	2 JX0095	glutathione transf
13	473.5	10.9	218	2 A39375	glutathione transf
14	470	10.8	218	2 A23732	glutathione transf
15	468.5	10.8	218	2 S13202	glutathione transf
16	467.5	10.7	218	2 B34159	glutathione transf
17	465.5	10.7	218	2 S01719	glutathione transf
18	463.5	10.7	218	2 A46048	glutathione transf
19	461.5	10.6	218	2 A29036	glutathione transf
20	460.5	10.6	218	2 B29231	glutathione transf
21	455.5	10.5	218	2 B09764	glutathione transf
22	454	10.4	218	2 A46143	glutathione transf
23	452.5	10.4	218	2 B28946	mu-class glutathio
24	447.5	10.3	225	2 A35295	glutathione transf
25	429.5	9.9	220	2 S18464	glutathione transf
26	385.5	8.9	219	2 S50146	glutathione transf
27	360	8.3	320	2 S55418	major allergen Dpl
28	306.5	7.0	322	2 E83974	UV-endonuclease ho
29	257.5	5.9	142	2 S17462	hypothetical prote
					glutathione transf

30	245.5	5.6	208	2 S41933	glutathione transf
31	244	5.6	210	2 S71958	glutathione transf
32	242	5.6	203	2 S13780	glutathione transf
33	240	5.5	210	1 A37378	glutathione transf
34	238	5.5	210	2 S71957	glutathione transf
35	237	5.4	210	1 XURTGP	glutathione transf
36	237	5.4	210	1 B55140	glutathione transf
37	236	5.4	210	2 I48112	glutathione S-tran
38	236	5.4	210	2 S71959	glutathione transf
39	235	5.4	210	2 JC6529	glutathione transf
40	233	5.4	209	1 A55140	glutathione transf
41	232	5.3	210	2 A41177	glutathione transf
42	230	5.3	210	2 A49180	glutathione transf
43	229.5	5.3	326	2 C75350	probable UV damage
44	222.5	5.1	223	1 A41031	glutathione transf
45	219.5	5.0	208	2 S03615	glutathione transf

ALIGNMENTS

RESULT 1

S71134
UV-endonuclease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 20-Jun-2000
C:Accession: S71134; T39815
R:Yasui, A.; Yamamoto, K.; Yamamoto, K.; Yasui, A.
Nucleic Acids Res. 24, 1267-1271, 1996
A:Title: Characterization of a UV-endonuclease gene from the fission yeast Schizosacch
A:Reference number: S71134; MUID:96188860
A:Accession: S71134
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-599 <TAK>
A:Cross-references: EMBL:D78571; NID:g1399000; PIDN:BAAL1415.1; PID:g1399001
A:Experimental source: strain Sp972
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Lucas, M.; Gaillardin, C.
submitted to the EMBL Data Library, June 1998
A:Reference number: 221861
A:Accession: T39815
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-599 <LYN>
A:Cross-references: EMBL:AL023859; PIDN:CAA19577.1; GSPDB:GN00067; SPDB:SPBC19C7.09c
A:Experimental source: strain 972h-; cosmid c19C7
C:Genetics:
A:Gene: UVDE; SPBC19C7.09c
A:Map position: 2

Query Match	71.6%	Score 3112;	DB 2;	Length 599;
Best Local Similarity	100.0%;	Pred. No. 1.8e-162;		
Matches 599;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	230	MLRLKRNIIQISKRIVFTILKQAFKGNHFCVPSVCTITYSRHFCLPDTLKLSPKSSKT	289	
Db	1	MLRLKRNIIQISKRIVFTILKQAFKGNHFCVPSVCTITYSRHFCLPDTLKLSPKSSKT	60	
QY	290	TLMLPQVNICANSFSAETPVDLKKNETELANISGPHKKSTSTSTRKRARSSKKKATDS	349	
Db	61	TLMLPQVNICANSFSAETPVDLKKNETELANISGPHKKSTSTSTRKRARSSKKKATDS	120	
QY	350	VSDKIDESVASYDSSTHLRSSRSKKKPVNYSSESESESEIQISKATKKVQKEEYVEE	409	
Db	121	VSDKIDESVASYDSSTHLRSSRSKKKPVNYSSESESESEIQISKATKKVQKEEYVEE	180	
QY	410	VDEKSLKNSSSDEFEPVPEQLPTISKRRRSKSAKLNLEKSTWNLDDHAPRMFDC	469	
Db	181	VDEKSLKNSSSDEFEPVPEQLPTISKRRRSKSAKLNLEKSTWNLDDHAPRMFDC	240	
QY	470	DKPTPWGRGLYACNLITLSMKERVFCSTRTITTIQRDGLSVKOLGTQNVLDLKL	529	

20Tokao

Tue Jan 15 14:12:06 2002

Db 241 DKPIPWGRGLYACLTITLRSKMKRERFCSTRTCTTTTQDGLSVKOLGTONVLDLIKLV 300
 QY 530 EWNHNFGLHFRVRSDDLPPFASHAKYGYTLEFAQSHLEEVKLANKNHRLTTHPGGYTQ 589
 Db 301 EWNHNFGLHFRVRSDDLPPFASHAKYGYTLEFAQSHLEEVKLANKNHRLTTHPGGYTQ 360
 QY 590 IASPREVVVDSAIRDLAYHDEILSRMKINEOLNKDAVLIILHGGTFEGKKTETLDRFRKNY 649
 Db 361 IASPREVVVDSAIRDLAYHDEILSRMKINEOLNKDAVLIILHGGTFEGKKTETLDRFRKNY 420
 QY 650 QRLSDSVKARLVLENDVSWSDLLPLCQELNIPVLVDWHHNIIVPCTLRREGSLDMLPL 709
 Db 421 QRLSDSVKARLVLENDVSWSDLLPLCQELNIPVLVDWHHNIIVPCTLRREGSLDMLPL 480
 QY 710 IPTIRETWRTKGITQKHYSADPTAISGMKRRASDRVDFPPCDPTMDLMTAEAKE 769
 Db 481 IPTIRETWRTKGITQKHYSADPTAISGMKRRASDRVDFPPCDPTMDLMTAEAKE 540
 QY 770 QAVFELCRRYELQNPCCPLEIMGPYDQTRDGYPPGAEKRLTARKRSRKEEVEDEK 828
 Db 541 QAVFELCRRYELQNPCCPLEIMGPYDQTRDGYPPGAEKRLTARKRSRKEEVEDEK 599

RESULT 2

A26484
 glutathione transferase (EC 2.5.1.18) - fluke (Schistosoma japonicum) (fragment)
 C:Species: Schistosoma japonicum
 C>Date: 25-Oct-1987 #sequence_revision 30-Sep-1989 #text_change 03-Feb-1994
 C:Accession: A94139; A26484; A28315
 R:Smith, D.B.; Davern, K.M.; Board, P.G.; Tiu, W.U.; Garcia, E.G.; Mitchell, G.F.
 Proc. Natl. Acad. Sci. U.S.A. 83, 8703-8707, 1986
 A:Title: Mr 26,000 antigen of Schistosoma japonicum recognized by resistant WEHI 129/J m
 A:Reference number: A94139; MUID:87041520
 A:Accession: A94139
 A:Molecule type: mRNA
 A:Residues: 1-219 <SMI>
 R:Smith, D.B.; Davern, K.M.; Board, P.G.; Tiu, W.U.; Garcia, E.G.; Mitchell, G.F.
 Proc. Natl. Acad. Sci. U.S.A. 84, 6541, 1987
 A:Reference number: A94181
 A:Contents: annotation; revision to residues 210-219
 C:Superfamily: glutathione transferase
 C:Keywords: transferase

Query Match 27.2%; Score 1184; DB 2; Length 219;
 Best Local Similarity 100.0%; Pred. No. 6.7e-58;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 TKLPILGYWKIKGLVOPTRLLLEYLEEYEEHLYERDEGDKWRNKKFELGLFPNLPYYI 61
 Db 1 TKLPILGYWKIKGLVOPTRLLLEYLEEYEEHLYERDEGDKWRNKKFELGLFPNLPYYI 60
 QY 62 DGDVKLTQSMATIRYADKNHNLGGCPKRAETSMLEGAVIDIRYGVSRVAYSKDPETLK 121
 Db 61 DGDVKLTQSMATIRYADKNHNLGGCPKRAETSMLEGAVIDIRYGVSRVAYSKDPETLK 120
 QY 122 VDFLSKLPMLKMFEDRLCHTKYINGDHVTHPDFMLYDALDVLVYMDPCLDAFPKLVCF 181
 Db 121 VDFLSKLPMLKMFEDRLCHTKYINGDHVTHPDFMLYDALDVLVYMDPCLDAFPKLVCF 180
 QY 182 KKREIATPQIDKYLKSSKYTAWPLQGWQATFGGDDHPK 220
 Db 181 KKREIATPQIDKYLKSSKYTAWPLQGWQATFGGDDHPK 219

RESULT 3

S55262
 uv-endonuclease - Neurospora crassa
 C:Species: Neurospora crassa
 C>Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
 C:Accession: S55262
 R:Yajima, H.; Takao, M.; Yasuhira, S.; Zhao, J.H.; Ishii, C.; Inoue, H.; Yasui, A.
 EMBO J. 14, 2393-2399, 1995

A:Title: A eukaryotic gene encoding an endonuclease that specifically repairs DNA dam
 A:Reference number: S55262; MUID:95292980
 A:Accession: S55262
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-656 <VAJ>
 A:Cross-references: EMBL:D11392; NID:g1526560; PID:g927215
 A:Genetics:
 A:Gene: uvel

Query Match 23.3%; Score 1011.5; DB 2; Length 656;
 Best Local Similarity 36.6%; Pred. No. 7.6e-48;
 Matches 241; Conservative 88; Mismatches 149; Indels 181; Gaps 18;

QY 339 ARSKKKAATDSVDKIDESVASYDSST-----HLRSSR-----SKKPVNNSSES- 386
 Db 3 SRKSKAALADTPOSESTFSSTLDSSAPSPARNLRKSGRNILQPSSEKDRDHEKRSGEEL 62
 QY 387 -----SERQISKATKKVKQKEEYVEEVEDEK 413
 Db 63 AGRMMGDANGCHLRECKBOEGVKMAIEGLARMERRLRQATKRQKKOLEED----- 114
 QY 414 SLANESSDEFEPVPEQLETPISKRRS-----RSAXNLEKESTMNLDHA----- 461
 Db 115 GIPVPSVVSFR-PTAPYHHKSTNAEEREAKPEVLKTHSKDVEREAIEGDDVVVKMEPAAT 173
 QY 462 ----PREMFCLDK-----PIWGRGLYACLTITLRSKMKRERFCSTRTCT 504
 Db 174 NIEPEDAQDAEAKGAPPAVNSVPLPWKGLRGYACLTILRNKSKPPIFSRTCRMA 233
 QY 505 TI-----ORDGLESVKOLGTONVLDLIKLVENHNFG 536
 Db 234 STVDHRHPLOFEDPEPHHLKKNPKDSKEPQDELGHKPVQELGLANARDIVKMLCWNKYG 293
 QY 537 IHFNRVSSDLFPFASHAKYGYTL-EFAQSHLEEVKLANKNHRLTTHPGGYTOIASPRE 595
 Db 294 IRLRLSSEMPFPASHVPHVGYKLAPFASEVLAAGRAVAELGHLRTHPGQFTQLGSPRK 353
 QY 596 VVDSATIRDLAYHDEILSRMKINEOLNKDAVLIILHGGTFEGKKTETLDRFRKNYQRLSDS 655
 Db 354 EVESAIRDLAYHDEILSRMKINEOLNKDAVLIILHGGTFEGKKTETLDRFRKNYQRLSDS 655
 QY 656 VKARLVLENDVSWSDLLPLCQELNIPVLVDWHHNIIVPCTLRREGSLDMLPL 711
 Db 414 CKNLVLVLENDVSWSDLLPLCQELNIPVLVDWHHNIIVPCTLRREGSLDMLPL 473
 QY 712 TIRETWRTKGITQKHYSADPTAISGMKRRASDRVDFPPCDPTMDLMTAEAKEQA 771
 Db 474 RIANTWRKKGKQKMHYSEPCD-CAVTPRHRKRPRVMTLPPCPDMDLMTAEAKEQA 532
 QY 772 VFELCRRYEL-----QNPPCP-----LEIMGPE--- 794
 Db 533 VFELMKTEKLPCEKINDMVPYDRDENRPPAPVKAPKKKGGKRRKRTTDEEAAPBEVD 592
 QY 795 --YQTRDG-----YYPGAPKRLTARKRSRKEEVEDEK 826
 Db 593 TAADDVKDAPEGPKPEVPEERAMGPNRVVWPLGCEENLKPKKREVKKGKVPVEDE 651

RESULT 4

A45556
 glutathione S-transferase - fluke (Schistosoma mansoni)
 C:Species: Schistosoma mansoni
 C>Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 03-May-1996
 C:Accession: A45556
 R:Wright, M.D.; Harrison, R.A.; Melder, A.M.; Newport, G.R.; Mitchell, G.F.
 Mol. Biochem. Parasitol. 49, 177-179, 1991
 A:Title: Another 26-kilodalton glutathione S-transferase of Schistosoma mansoni.
 A:Reference number: A45556; MUID:92131046
 A:Accession: A45556
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid

QY	5	PILGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLFEPNLPYYIDGD	64
Db	3	PKFEGYWKVGLVQPTRLLLEHLEETVEERAYDRNEIDAWSNDFKLGLEFPNLPYYIDGD	62
QY	65	VKLQTSMAIIRYIADKHNLGLGCPKERAERISMLEGAVLDIRYGVSIAYSKDFETLKVDFF	124
Db	63	FKLTQSMARIYIADKHNLGLGCPKERAERISMLEGAVLDIRMGVLRIRAYNKYEETLKVDFF	122
QY	125	LSKLPEMLKMFEDRLCHTKTYLNGDHWTHPDFDMLYDALDWLYMDPCLDAPFKLVCFPKR	184
Db	123	LNKLPGRLKMFEDRLSNKTYLNGCVTHPDFDMLYDALDWLYMDSOCLNEFPKLVSFKKC	182
QY	185	IEAIPQIDRYLKSSXYIAWPLOGWOATFGGDHPKK	220
Db	183	IEDLPQIKNLNSRXYIKWPLOGWDATFGGDTTPK	218

A:Accession: GB:U13448
A:Cross-references: GB:J04632; NID:gl9547; PIDN:AAA37705.1; PID:g309260
A:Note: the authors translated the codon GAG for residue 49 as Gly and GAG for residue
R:Townsend, A.J.; Goldsmith, M.E.; Pickett, C.B.; Cowan, K.H.
J. Biol. Chem. 264, 21582-21590, 1989
A:Title: Isolation, characterization, and expression in *Escherichia coli* of two murin
A:Reference number: A34159; MUID:90094327
A:Accession: A34159
A:Molecule type: mRNA
A:Residues: 1-218 <TOW>
A:Cross-references: GB:J04632; NID:gl9547; PIDN:AAA37705.1; PID:g309260
R:Pearson, W.R.; Reinhart, J.; Sisk, S.C.; Anderson, K.S.; Adler, P.N.
J. Biol. Chem. 263, 13324-13332, 1988
A:Title: Tissue-specific induction of murine glutathione transferase mRNAs by butylat
A:Reference number: A92668; MUID:88330838
A:Contents: clone pGT 875
A:Accession: A28946
A:Molecule type: mRNA

A:Residues: 2-218 <PEA>
A:Cross-references: GB:J03952; NID:gl93687; PIDN:AAA37747.1; PID:g309278
R:Pearson, W.R.; Windle, J.J.; Morrow, J.F.; Benson, A.M.; Talalay, P.
J. Biol. Chem. 258, 2052-2062, 1983
A:Title: Increased synthesis of glutathione S-transferases in response to anticarcinogen
A:Reference number: A92411; MUID:83109018
A:Accession: A20831
A:Molecule type: protein
A:Residues: 2-7, 'X', 8-13, 'X', 15-23, 'X', 25, 'XX', 28, 'N', 30-31, 'X', 33, 'X', 35-39, 'M', 41 <PEI>
R:Manervik, B.; Alin, P.; Guthenberg, C.; Jonsson, H.; Tahir, M.K.; Warholm, M.; Jorva
Proc. Natl. Acad. Sci. U.S.A. 82, 7202-7206, 1985
A:Title: Identification of three classes of cytosolic glutathione transferase common to
A:Reference number: A24735; MUID:86042634
A:Accession: A24735
A:Molecule type: protein
A:Residues: 2-7, 'X', 8-13, 'X', 15-23, 'X', 25 <MAN>
R:Hayes, J.D.; Kerr, L.A.; Peacock, S.D.; Cronshaw, A.D.; McLellan, L.I.
Biochem. J. 277, 501-512, 1991
A:Title: Hepatic glutathione S-transferases in mice fed on a diet containing the anticarcinogen
A:Reference number: S16933; MUID:91315425
A:Accession: S16933
A:Molecule type: protein
A:Residues: 2-36 <HAY>
R:Fernandes, C.L.; Dong, J.H.; Roebuck, B.D.; Chisari, F.V.; Montali, J.A.; Schmidt Jr.
Arch. Biochem. Biophys. 331, 104-116, 1996
A:Title: Elevations of hepatic quinone reductase, glutathione, and alpha- and mu-class
A:Reference number: S71311; MUID:96268466
A:Accession: S71312
A:Molecule type: protein
A:Residues: 97-108 <FER>
C:Genetics: GSTM1
A:Introns: 12/3; 38/1; 59/3; 87/1; 120/3; 152/3; 189/3
C:Superfamily: glutathione transferase
C:Keywords: dimer; transferase
F:2-218/Product: glutathione transferase mul #status experimental <MAT>

Query Match 11.1%; Score 481.5; DB 2; Length 218;
Best Local Similarity 44.4%; Pred. No. 1.4e-19;
Matches 91; Conservative 38; Mismatches 71; Indels 5; Gaps 1;

QY 6 ILGWYKTKGLVQPTRLLEYLEEKEVEHLYERDEG-----DKWRNKKFELGLEFPNLPYY 60
Db 4 ILGWYVNGLTHPIRLLEYLETDSYDEKRYTGDAPDPSQWLNKEKFKGLDFPNLPYL 63
QY 61 IDGDKVLTQSMATIRYADKHNMMLGGCPKERAELSMLEGAVIDIRYGVSRAYSXDFETL 120
Db 64 IDGSHKITQSNALIRYLARKHHLGDETEERIRADIVENQVMDTRMQILMLCYNPDFEKQ 123
QY 121 KVDLSKLPKMLKMFEDRLCHKTYLNGDHVTHPDFDMLYDALDVLVMDPCLDAFPKLVK 180
Db 124 KPEFLKTIPEKMLKLYSEFLGKRFWAGDKVTVYDFLAYDILQYRMFEPKCLDAFPNLRD 183
QY 181 FKKRIEAIPTQIDKYLKSSKYIAMPL 205
Db 184 FLAREGLKKISAYMKSSRYIATPI 208

RESULT 8
A29794
glutathione transferase (EC 2.5.1.18) class mu chain 3 - rat
N:Alternate names: glutathione S-transferase Yb1; ligandin
C:Species: Rattus norvegicus (Norway rat)
C:Date: 28-Aug-1989 #sequence.revision 28-Aug-1989 #text change 18-Jun-1999
A:Accession: A29794; A25510; A24085; B61004; P24735; A33397; S27111; S17167; A26187
R:Chang, C.; Saltzman, A.G.; Sorensen, N.S.; Hilpakka, R.A.; Liao, S.
J. Biol. Chem. 262, 11901-11903, 1987
A:Title: Identification of glutathione S-transferase Yb-1 mRNA as the androgen-repressed
A:Reference number: A29794; MUID:87308179
A:Accession: A29794
A:Molecule type: mRNA

A:Residues: 1-218 <CHA>
A:Cross-references: GB:J02810; NID:g204514; PIDN:AAA41293.1; PID:g204515
R:Lai, H.C.J.; Grove, G.; Tu, C.P.D.
Nucleic Acids Res. 14, 6101-6114, 1986
A:Title: Cloning and sequence analysis of a cDNA for a rat liver glutathione S-transfer
A:Reference number: A25510; MUID:86312882
A:Accession: A25510
A:Molecule type: mRNA
A:Residues: 1-218 <LA1>
A:Cross-references: GB:X04229; NID:g56337; PIDN:CAA27811.1; PID:g56338
R:Ding, G.J.F.; Lu, A.Y.H.; Pickett, C.B.
J. Biol. Chem. 260, 13268-13271, 1985
A:Reference number: A24085; MUID:86033768
A:Accession: A24085
A:Molecule type: mRNA
A:Residues: 1-198, 'NC', 201-218 <DIN>
A:Cross-references: GB:M11719; NID:g204502; PIDN:AAA41287.1; PID:g204503
A:Experimental source: clone pGTA/C44
R:Chang, L.H.; Hsieh, J.C.; Chen, W.L.; Tam, M.F.
Electrophoresis 11, 589-593, 1990
A:Title: Identification of rat liver glutathione S-transferase Yb subunits by partial
A:Reference number: A61004; MUID:91031411
A:Accession: A61004
A:Molecule type: protein
A:Residues: 1-26 <CH2>
R:Manervik, B.; Alin, P.; Guthenberg, C.; Jonsson, H.; Tahir, M.K.; Warholm, M.; Jor
Proc. Natl. Acad. Sci. U.S.A. 82, 7202-7206, 1985
A:Title: Identification of three classes of cytosolic glutathione transferase common
A:Reference number: A24735; MUID:86042634
A:Accession: F24735
A:Molecule type: protein
A:Residues: 2-20 <MAN>
R:Hsieh, J.C.; Liu, L.F.; Chen, W.L.; Tam, M.F.
Biochem. Biophys. Res. Commun. 162, 1147-1154, 1989
A:Title: Expression of Yb-1 glutathione S-transferase using a baculovirus expression
A:Reference number: A33397; MUID:89350924
A:Accession: A33397
A:Molecule type: protein
A:Residues: 2-21; 212-218 <HS1>
R:Katusz, R.M.; Bono, B.; Colman, R.F.
Arch. Biochem. Biophys. 298, 667-677, 1992
A:Title: Identification of Yb(115) labeled by S-(4-bromo-2,3-dioxobutyl)glutathione
A:Reference number: S27111; MUID:93037509
A:Accession: S27111
A:Molecule type: protein
A:Residues: 2-20; 83-86, 'X', 88-96; 109-115, 'X', 117-122 <KAT>
R:Hsieh, J.C.; Huang, S.C.; Chen, W.L.; Lai, Y.C.; Tam, M.F.
Biochem. J. 278, 293-297, 1991
A:Title: Cysteine-86 is not needed for the enzymic activity of glutathione S-transfer
A:Reference number: S17167; MUID:91354218
A:Accession: S17167
A>Status: preliminary
A:Molecule type: protein
A:Residues: 2-218 <HS2>
C:Superfamily: glutathione transferase
C:Keywords: dimer; transferase

Query Match 11.1%; Score 481.5; DB 2; Length 218;
Best Local Similarity 44.4%; Pred. No. 1.4e-19;
Matches 91; Conservative 38; Mismatches 71; Indels 5; Gaps 1;

QY 6 ILGWYKTKGLVQPTRLLEYLEEKEVEHLYERDEG-----DKWRNKKFELGLEFPNLPYY 60
Db 4 ILGWYVNGLTHPIRLLEYLETDSYDEKRYTGDAPDPSQWLNKEKFKGLDFPNLPYL 63
QY 61 IDGDKVLTQSMATIRYADKHNMMLGGCPKERAELSMLEGAVIDIRYGVSRAYSXDFETL 120
Db 64 IDGSHKITQSNALIRYLARKHHLGDETEERIRADIVENQVMDTRMQILMLCYNPDFEKQ 123
QY 121 KVDLSKLPKMLKMFEDRLCHKTYLNGDHVTHPDFDMLYDALDVLVMDPCLDAFPKLVK 180
Db 124 KPEFLKTIPEKMLKLYSEFLGKRFWAGDKVTVYDFLAYDILQYRMFEPKCLDAFPNLRD 183
QY 181 FKKRIEAIPTQIDKYLKSSKYIAMPL 205
Db 184 FLAREGLKKISAYMKSSRYIATPI 208

Db 124 KPEFLKTIPEKMKLYSEFLGKRPWFGAGDKVTVYVDFLAYDILDQYHIEPKCLDAFPNPKD 183
 QY 181 FKKRTAIPQIDKYLKSKSYIAWPL 205
 Db 184 FLARFEGGLKISAYMKSSRYLSTPI 208

RESULT 9
 S32425
 glutathione transferase (EC 2.5.1.18) class mu, GSTM4 (version 2) - human
 N:Alternate names: glutathione transferase mu4
 C:Species: Homo sapiens (man)
 C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 18-Jun-1999
 C:Accession: S32425; S29337
 R:Zhong, S.; Spurr, N.K.; Hayes, J.D.; Wolf, C.R.
 Biochem. J. 291, 41-50, 1993
 A:Title: Deduced amino acid sequence, gene structure and chromosomal location of a novel
 A:Reference number: S32424; MUID:93228631
 A:Accession: S32425
 A:Molecule type: DNA
 A:Residues: 1-218 <2HO>
 A:Cross-references: EMBL:X68677; NID:g31934; PIDN:CAA48637.1; PID:g31935
 A:Note: The authors translated the codon AAG for residue 182 as Arg and CCA for residue
 C:Genetics:
 A:Gene: GDB:GSTM4
 A:Cross-references: GDB:134191; OMIM:138333
 A:Map position: lp13.3-1p13.3
 A:Introns: 12/3; 38/1; 59/3; 87/1; 120/3; 152/3; 189/3
 C:Superfamily: glutathione transferase
 C:Keywords: dimer; transferase

Query Match 11.0%; Score 478.5; DB 2; Length 218;
 Best Local Similarity 44.1%; Pred. No. 2e-19;
 Matches 94; Conservative 37; Mismatches 77; Indels 5; Gaps 1;
 QY 6 ILGYWIKGLVQPTRLLEYLEEYEEHLYERD-----ECDKWRNKKFELGLEFNPPLY 60
 Db 4 ILGYWIRGLAHAIIRLLEVTSSYEKKYTMGGAPDYDRSQWLNEKPKLGLDFNPPLY 63
 QY 61 IDGDVTKLTQSMATIRYIADKHNMGLGCPKERAIEISMLEGAVLDIRYGVSRVAYSKDFETPL 120
 Db 64 IDGAHTQSNALICVIAKHNLCJGTEEEKIRVDILENQAMDVSNQLARVCYSPDFEKL 123
 QY 121 KVDFLSKLPEMLKMFEDRLCHKTYLNGDHTVHPDFMLYDALDVLYMDPMCLDAPFKLVC 180
 Db 124 KPEYLEELPTMQHFSQFLGKRPWFVGDKITFVDFLAYDVLHLRIFEPNCLDAPFNPKD 183
 QY 181 FKKRIEAIPOIDKYLKSSKYIAWPLQGWATFG 213
 Db 184 FTSRFEGLKISAYMKSSRFLPKPLYTRMAVWG 216

RESULT 10
 S65674
 glutathione transferase (EC 2.5.1.18) class mu chain 7.8 - rabbit
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Jun-1999
 C:Accession: S65674; S30380
 R:Lee, S.H.; Lee, Y.S.; Han, J.S.; Kim, Y.S.; Koh, J.K.
 Arch. Biochem. Biophys. 318, 424-429, 1995
 A:Title: Cloning and expression of a cDNA for Mu-Class Glutathione S-transferase from ra
 A:Reference number: S65674; MUID:95251394
 A:Accession: S65674
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-218 <LEP>
 A:Cross-references: EMBL:L23766; NID:g388198; PIDN:AAA69665.1; PID:g388199
 R:Primiano, T.; Novak, R.F.
 Arch. Biochem. Biophys. 301, 404-410, 1993
 A:Title: Purification and characterization of class mu glutathione S-transferase isozyme
 A:Reference number: S30380; MUID:93213177
 A:Accession: S30380

A:Molecule type: protein
 A:Residues: 2-21 <PRI>
 C:Superfamily: glutathione transferase
 C:Keywords: dimer; transferase

Query Match 11.0%; Score 477.5; DB 2; Length 218;
 Best Local Similarity 44.6%; Pred. No. 2.3e-19;
 Matches 95; Conservative 36; Mismatches 77; Indels 5; Gaps 1;
 QY 7 LGVWKIKGLVQPTRLLEYLEEYEEHLYERDEG-----DKWRNKKFELGLEFNPPLY 61
 Db 5 LGTWDVRGLALPIRMILEYTDTSYEKKYTMGDAPNYDQSKWLSEKFTLGLDFNPPLY 64
 QY 62 DGDVTKLTQSMATIRYIADKHNMGLGCPKERAIEISMLEGAVLDIRYGVSRVAYSKDFETLK 121
 Db 65 DGTHTKLTQSNALIRYLARKHGLCGTEEEKIRVDILENQMDNRFLQVNVVCSYSPDFEKL 124
 QY 122 VDFLSKLPKMFEDRLCHKTYLNGDHTVHPDFMLYDALDVLYMDPMCLDAPFKLVC 181
 Db 125 PEYLKGLPEKQLQYQFLGSLFWFAGDKITFADFVLVYDVLQDNRIFFVPGCLDAPFNPKD 184
 QY 182 KKRRIEAIPOIDKYLKSSKYIAWPLQGWATFG 214
 Db 185 HVRFEGLPKISAYMKSSRFRIRVPVFLKKTATWG 217

RESULT 11
 A47486
 glutathione transferase (EC 2.5.1.18) class mu, GSTM4 (version 1) - human
 C:Species: Homo sapiens (man)
 C>Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 18-Jun-1999
 C:Accession: A47486; B47486; S36782; I37438; S45685
 R:Comstock, K.E.; Johnson, K.J.; Rifkenbery, D.; Henner, W.D.
 J. Biol. Chem. 268, 16958-16965, 1993
 A:Title: Isolation and analysis of the gene and cDNA for a human mu class glutathione
 A:Reference number: A47486; MUID:93352467
 A:Accession: A47486
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-218 <COM>
 A:Cross-references: GB:M96233
 A:Accession: B47486
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-218 <CO2>
 A:Cross-references: GB:M96234; NID:g306818; PIDN:AAA57347.1; PID:g306819
 A:Experimental source: Hela cells
 R:Ross, V.L.; Board, P.G.
 Biochem. J. 294, 373-380, 1993

A:Title: Molecular cloning and heterologous expression of an alternatively spliced hu
 A:Reference number: S36782; MUID:93384505
 A:Accession: S36782
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-218 <ROS>
 A:Cross-references: EMBL:M99422
 A:Experimental source: testis
 R:Comstock, K.E.; Widersten, M.; Hao, X.Y.; Henner, W.D.; Mannervik, B.
 Arch. Biochem. Biophys. 311, 487-495, 1994
 A:Title: A comparison of the enzymatic and physicochemical properties of human glutat
 A:Reference number: S45684; MUID:94263230
 A:Accession: I37438
 A:Contents: annotation
 R:Taylor, J.B.; Oliver, J.; Sherrington, R.; Pemble, S.E.
 Biochem. J. 274, 587-593, 1991
 A:Title: Structure of human glutathione S-transferase class Mu genes.
 A:Reference number: I37438; MUID:91174774
 A:Accession: I37438
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 39-120 <RES>
 A:Cross-references: EMBL:X56837; NID:g31936; PIDN:CAA40167.1; PID:g31937
 C:Genetics:

Db 185 ISRFEGLEKISAYMKSSRFLPRPVTKMAVWG 216

RESULT 14

A23732

glutathione transferase (EC 2.5.1.18) mu - golden hamster

C:Species: Mesocricetus auratus (golden hamster)

C:Date: 30-Dec-1991 #sequence_revision 30-Dec-1991 #text_change 18-Jun-1999

C:Accession: A23732; S21908

R:Norris, J.S.; Schwartz, D.A.; MacLeod, S.L.; Fan, W.; O'Brien, T.J.; Harris, S.E.; T.

Mol. Endocrinol. 5, 979-986, 1991

A:Title: Cloning of a mu-class glutathione S-transferase complementary DNA and character

A:Reference number: A23732; MUID:92049380

A:Accession: A23732

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-218 <NOR>

A:Cross-references: EMBL:X61033; NID:g49638; PIDN:CAA43368.1; PID:g49639

A:Note: submitted to the EMBL Data Library, July 1991

C:Superfamily: glutathione transferase

C:Keywords: transferase

Query Match

Best Local Similarity 10.8%; Score 470; DB 2; Length 218;

Matches 90; Conservative 44.3%; Pred. No. 5.9e-19;

Mismatches 39; Gaps 2; Indels 68; Gaps 2;

Qy 4 LPI-LGYWKIRKGLVQPTRLLEYLEEKYEHLVERDEG-----DKWRNKKFELGLEFPNL 57

Db 1 MPVTGLGYWDIRGLAHAIRLLLEYDDTSYEKKYTGADPNFDRSQWLNEKFKLGDFPNL 60

Qy 58 PYYIDGVKLTQSWAIRIYADKNHMLGGCGPKRAEISMLEGAVLDIRYGVSRISYKDF 117

Db 61 PYLIDGSHKITQSNAILRYIARKHDLGCGTEERIQLDLENQAMQTRMQLAMVCYSPDF 120

Qy 118 ETLKVDFLSKLPEMKWPFEDRLCHTKTYLNGDHVTHPDPMYDLDVLYMDPMCLDAFPK 177

Db 121 EKRFYLEGPEKMKLYSEFLGRKSWFAGDKITYVDVFLYDLDQHRIFAPKCLDAFPN 180

Qy 178 LVCFKKRIEAIPOIDKYLKSSKY 200

Db 181 LKDFLARFEGLEKISAYMKSSRF 203

RESULT 15

S13202

glutathione transferase (EC 2.5.1.18) Y1 - Chinese hamster

C:Species: Cricetulus griseus (Chinese hamster)

C:Date: 18-Feb-1994 #sequence_revision 06-Jan-1995 #text_change 18-Jun-1999

C:Accession: S13202

R:Robert de Saint Vincent, B.; Hyrien, O.; Debatisse, M.; Buttin, G.

Eur. J. Biochem. 193, 19-24, 1990

A:Title: Coamplification of mu class glutathione S-transferase genes and an adenylate de

A:Reference number: S13202; MUID:91031445

A:Accession: S13202

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-218 <ROB>

A:Cross-references: GB:X57489; NID:g49541; PIDN:CAA40726.1; PID:g49542

A:Note: in Genbank entry Clv1, release 109, the source is designated as Cricetulus longi

C:Superfamily: glutathione transferase

C:Keywords: dimer; transferase

Query Match

Best Local Similarity 10.8%; Score 468.5; DB 2; Length 218;

Matches 89; Conservative 43.4%; Pred. No. 7.2e-19;

Mismatches 36; Gaps 1; Indels 75; Gaps 1;

Qy 6 ILGYWKIRKGLVQPTRLLEYLEEKYEHLVERDEG-----DKWRNKKFELGLEFPNL 60

Db 4 ILGYWNRGLTNPTRLLEVTDSSEYEEKVTMGDAPDSQRSQWLNEKFKLGDFPNL 63

Qy 61 IDGDKLTQSWAIRIYADKNHMLGGCGPKRAEISMLEGAVLDIRYGVSRISYKDFETL 120

Db 64 IDGSHKITQSNAILRYIARKHNLGCGTEERIRVDIVENQAMDTRMQLIMLCYNPDFEQ 123

Qy 121 KVDFLSKLPEMKWPFEDRLCHTKTYLNCDDHVTHTDPMYDLDVLYMDPMCLDAFPKLV 180

Db 124 KPFLTKTIPKMKMYSEFLGRKSWFAGDKVTLGCGFLAYDVLQYQMPKCLDPPFNLD 183

Qy 181 FKRIEAIPOIDKYLKSSKYIAWPL 205

Db 184 FLARFEGLEKISAYMKTSRFLRPI 208

Search completed: January 15, 2002, 13:52:18

Job time: 387 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:06:15 ; Search time 134.53 Seconds
(without alignments)
713.258 Million cell updates/sec

Title: US-09-724-296-36

Perfect score: 3482

Sequence: 1 MPSRKSAAALDTQSEST.....REVKKGVPEVEDEGEFDG 656

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3482	100.0	656	Q01408	Q01408 neurospora
2	1011.5	29.0	599	Q10988	Q10988 schizosacch
3	377	10.8	322	Q9K9P8	Q9K9P8 bacillus ha
4	359	10.3	317	Q9L4F0	Q9L4F0 bacillus ce
5	359	10.3	325	Q9L4E9	Q9L4E9 bacillus ce
6	261.5	7.5	305	Q9S0M9	Q9S0M9 deinococcus
7	261.5	7.5	326	Q9RTE6	Q9RTE6 deinococcus
8	160.5	4.6	1280	Q75158	Q75158 homo sapien
9	160.5	4.6	1341	Q9UKV3	Q9UKV3 homo sapien
10	155.5	4.5	5327	Q76891	Q76891 drosophila
11	152.5	4.4	1763	Q9JJKX5	Q9JJKX5 mus musculu
12	144	4.1	16215	Q9NFS3	Q9NFS3 drosophila
13	143.5	4.1	828	Q9VEU8	Q9VEU8 drosophila
14	143.5	4.1	914	Q9H945	Q9H945 homo sapien
15	141.5	4.1	1110	Q91255	Q91255 petromyzon
16	141	4.0	776	Q9H1R4	Q9H1R4 homo sapien
17	140.5	4.0	1338	Q9JIX8	Q9JIX8 mus musculu
18	140.5	4.0	3263	Q917U3	Q917U3 drosophila
19	140.5	4.0	6815	Q917U4	Q917U4 drosophila

20	139.5	4.0	1510	5	061802	061802 caenorhabdi
21	139	4.0	1815	2	Q9K0B4	Q9K0B4 neisseria m
22	138.5	4.0	1520	4	015087	015087 homo sapien
23	138.5	4.0	1781	4	Q9UKX0	Q9UKX0 homo sapien
24	138	4.0	1547	5	Q26471	Q26471 schistocerc
25	137.5	3.9	806	10	Q9M8T5	Q9M8T5 arabidopsis
26	137.5	3.9	964	5	Q9VKD8	Q9VKD8 drosophila
27	137.5	3.9	1197	3	Q9C3Y7	Q9C3Y7 candida alb
28	137	3.9	735	5	Q9VTN4	Q9VTN4 drosophila
29	137	3.9	785	5	Q9GQ82	Q9GQ82 drosophila
30	136.5	3.9	1591	11	P97868	P97868 mus musculu
31	136	3.9	3261	4	Q9V556	Q9V556 homo sapien
32	136	3.9	3942	11	088737	088737 mus musculu
33	135.5	3.9	971	5	045785	045785 caenorhabdi
34	135	3.9	767	5	Q9U234	Q9U234 caenorhabdi
35	135	3.9	1530	11	Q9EQZ7	Q9EQZ7 mus musculu
36	134.5	3.9	837	5	Q9VYV7	Q9VYV7 drosophila
37	134.5	3.9	1695	5	Q9NK53	Q9NK53 drosophila
38	134.5	3.9	1711	5	Q9VJL0	Q9VJL0 drosophila
39	134	3.8	3576	11	Q9QZK2	Q9QZK2 mus musculu
40	133.5	3.8	583	4	Q9UKV2	Q9UKV2 homo sapien
41	133.5	3.8	980	5	Q96246	Q96246 plasmodium
42	133.5	3.8	1399	11	Q9JIS0	Q9JIS0 rattus norv
43	133	3.8	718	13	073619	073619 xenopus lae
44	133	3.8	795	5	Q9V8K6	Q9V8K6 drosophila
45	133	3.8	2073	4	Q9UKW2	Q9UKW2 homo sapien

ALIGNMENTS

RESULT 1
Q01408
ID Q01408; PRELIMINARY: PRT: 656 AA.
AC Q01408;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE UV-ENDORNUCLEASE.
GN UVEL.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=95292980; PubMed=7774597;
RA Yajima H., Takao M., Yasuhira S., Zhao J.H., Ishii C., Inoue H.,
RA Yasui A.;
RT "A eukaryotic gene encoding an endonuclease that specifically repairs
RT DNA damaged by ultraviolet light.";
RL EMBO J. 14:2393-2399(1995).
DR EMBL; D11392; CAB20113.1; .
KW Endonuclease.
SQ SEQUENCE 656 AA; 74413 MW; 4B231E7A8735C86F CRC64;

Query Match 100.0%; Score 3482; DB 3; Length 656;
Best Local Similarity 100.0%; Pred. No. 4.4e-227;
Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPSRKSAAALDTQSESTFSTLDSAPSARNLRSGRNILQPSSEKDRDHEKRSGE 60
Db 1 MPSRKSAAALDTQSESTFSTLDSAPSARNLRSGRNILQPSSEKDRDHEKRSGE 60
Qy 61 ELAGRMGKDGANGHCLREGKEQECVKMAIEGLARMPRIQORATKROKQLEEDGIPVPS 120
Db 61 ELAGRMGKDGANGHCLREGKEQECVKMAIEGLARMPRIQORATKROKQLEEDGIPVPS 120
Qy 121 VVSREPTAPYHHKSTNAEREAKEPVLKTHSKDVEREAETGVDDVVKMEPAATNIIEPD 180
Db 121 VVSREPTAPYHHKSTNAEREAKEPVLKTHSKDVEREAETGVDDVVKMEPAATNIIEPD 180

```

QY 181 AODAAERGAARPPAVNSYLLPLPWKGRIGYACLTNTYLRNAKPPPIFSRTCRMASIVDHRH 240
DB 181 AODAAERGAARPPAVNSYLLPLPWKGRIGYACLTNTYLRNAKPPPIFSRTCRMASIVDHRH 240
QY 241 PLOFDEPEHHLKNNPKDSKEPQDELGHKGFVOELGLANARDIVKMLCWNENEKYGIRFLRLS 300
DB 241 PLOFDEPEHHLKNNPKDSKEPQDELGHKGFVOELGLANARDIVKMLCWNENEKYGIRFLRLS 300
QY 301 SEMEPFASHPVHGKYLAPFASVLAELAGRAVAELGHLRLLTHPGQFTQLGSPKRVESAI 360
DB 301 SEMEPFASHPVHGKYLAPFASVLAELAGRAVAELGHLRLLTHPGQFTQLGSPKRVESAI 360
QY 361 ROLYHDELLSLLKLPEQONRDVMI IHMGQFGDKAATLERFKRNYARLSOSCKNRLVL 420
DB 361 ROLYHDELLSLLKLPEQONRDVMI IHMGQFGDKAATLERFKRNYARLSOSCKNRLVL 420
QY 421 ENDDVGVTHDILLPYCEELNIPMVDYHHNHCIFDPAHLREGTILDISPKLOERIANTWK 480
DB 421 ENDDVGVTHDILLPYCEELNIPMVDYHHNHCIFDPAHLREGTILDISPKLOERIANTWK 480
QY 481 RKGIRKQKMHYSEPCGAVTPDRRKRHRPRVMTLPCCPPDMOLMIEAKDKQAVFELMRTF 540
DB 481 RKGIRKQKMHYSEPCGAVTPDRRKRHRPRVMTLPCCPPDMOLMIEAKDKQAVFELMRTF 540
QY 541 KLPGFEKINDMVPYDRDDENRPPVAPKPKKGGKRRKTTDEEAAPEEVDTAADDVYKD 600
DB 541 KLPGFEKINDMVPYDRDDENRPPVAPKPKKGGKRRKTTDEEAAPEEVDTAADDVYKD 600
QY 601 APEGKPEVEERAMGGPNYRVWPLGCEWLKPKKREVKKGKVPVEEDGEGFDG 656
DB 601 APEGKPEVEERAMGGPNYRVWPLGCEWLKPKKREVKKGKVPVEEDGEGFDG 656

RESULT 2
Q10988 PRELIMINARY; PRT; 599 AA.
AC Q10988; P87339;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE UV-ENDORNUCLEASE.
GN UVDE OR UVEL+.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sp972;
RX MEDLINE=96188860; PubMed=8614629;
RA Takao M., Yonemasu R., Yamamoto K., Yasui A.;
RT "Characterization of a uv endonuclease gene from the fission yeast
RT Schizosaccharomyces pombe and its bacterial homolog.";
RL Nucleic Acids Res. 24:1267-1271(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97175806; PubMed=9023111;
RA Davey S., Nass M.L., Ferrer J.V., Sidik K., Eisenberger A.,
RA Mitchell D.L., Freyer G.A.;
RT "The fission yeast UVDR DNA repair pathway is inducible.";
RL Nucleic Acids Res. 25:1002-1008(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Lyne M., Rajandream M.A., Barrell B.G., Lucas M., Gaillardin C.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; D78571; BAAL1415.1;
DR EMBL; U78487; AAC49664.1;
DR EMBL; AL023859; CAA19577.1;
KW Endonuclease.
SQ SEQUENCE 599 AA; 68816 MW; C09ED7219D5895BE CRC64;

```

```

Query Match 29.0%; Score 1011.5; DB 3; Length 599;
Best Local Similarity 36.6%; Pred. No. 2.5e-60;
Matches 241; Conservative 88; Mismatches 149; Indels 181; Gaps 18;

QY 3 SRKSKAAALDTQSESSSTFSSLTDSAPSPARNLRSGNRILQPSSEKDRDHEKRSGEEL 62
DB 110 ARSSKKATDSVSDKIDESVYDSST-----HLRSSR-----SKPVNYSSESE- 157
QY 63 AGRMMKDGANGHCLRGKEQEEGVKMAIEGLARMERLQORATKROKKQLEED----- 114
DB 158 -----SEQISKATKKVKQKEEYVEVDK 184
QY 115 GIPVPVWSRF-PTAPYHHKSTNAEREAKEPVLTHTSKDSDVEREAEIGVDDVVKMEPAAT 173
DB 185 SLANESSDEFEVPEQLETPISKRRS-----RSSAKNLEKESIMNLDHA----- 232
QY 174 NIIEPEDAADAERGAARPPAVNSYLLPLPWKGRIGYACLTNTYLRNAKPPPIFSRTCRMA 233
DB 233 ---PREMPDCLDK-----PIWGRGLGYACLTILRSMKERVFCSTCRIT 275
QY 234 SIYDHRHPLQFDEPEBHLKNNPKDSKEPQDELGHKGFVOELGLANARDIVKMLCWNENEKG 293
DB 276 TI-----QRDGLSVKQLGTONVLDLILKLVENHNFG 307
QY 294 IRELRLSSMFPFASHPVHGKYLAPFASVLAELAGRAVAELGHLRLLTHPGQFTQLGSPRK 353
DB 308 IHEMRYSSDLFPFASHAKYGYTL-EFAQSHLBEVGVGLANKYHNRLTMHPGQVYTIASPRE 366
QY 354 EYVESAIRDLEYHDELLSLLKLPEQONRDVMI IHMGQFGDKAATLERFKRNYARLSOS 413
DB 367 VVDSAIRDLAYHDEILSRMKLNQNLKDAVLI IHUGGTFEGKKETLDRPKKNYQRLSOS 426
QY 414 CKNRVLNDDVGVTHDILLPYCEELNIPMVDYHHNHCIFDPAHLREGTILDISPKLOE 473
DB 427 VKARLVLEDDVSVSQDILLPLQELNIPLYLDWHHNIV--PGTLREGSIDLM--PLIP 482
QY 474 RIANTWRKKGKIKOMHYSEPCD-GAVTPDRRKRHRPRVMTLPCCPPDMOLMIEAKDKQEA 532
DB 483 TIRETWTRKGITQKHYSASADPTAISGMKRRRAHSRDFDFPCDPTMDLMIEAKDKQEA 542
QY 533 VFELMRTFKLPGFEKINDMVPYDRDDENRPPVAPKPKKGGKRRKTTDEEAAPEEVD 592
DB 543 VFELCRRYEL-----QNPPCP-----LEIMGPE--- 565
QY 593 TAADDVKADEPGKPEVEERAMGGPNYRVWPLGCEWLKPKKREVKKGKVPVEEVEDE 651
DB 566 --YDQTRDG-----YPPGAEKRLTARKRSRK-----EEVEED 597

```

```

RESULT 3
Q9K9P8 PRELIMINARY; PRT; 322 AA.
AC Q9K9P8
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE BH2597 PROTEIN.
GN BH2597.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";

```



```
RESULT 10
076891 ID 076891 PRELIMINARY; PRT: 5327 AA.
AC 076891
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE EG:49B4.1 PROTEIN.
GN FUTSCH OR EG:49B4.1 OR CG3064.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Benos P.;
RA Papagiannakis G., Spanos L., Siden-Kiamos I., Louis C.;
RT "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
RA Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL: AL031128; CAA20006.1; -.
DR FlyBase: FBgn0015390; futsch.
SQ SEQUENCE 5327 AA; 575942 MW; FFEER23A118FF38A CRC64;

Query Match 4.5%; Score 155.5; DB 5; Length 5327;
Best Local Similarity 19.4%; Pred. No. 0.28;
Matches 151; Conservative 104; Mismatches 306; Indels 219; Gaps 34;

QY 3 SRKSKAAALDTPQSFSTFS-----STLSSAPSPARNLRRSGRNLIQPSSE-KORDHE 55
DB 1378 TEKSPLASKETSRPSAFSGVKEDTEQTKKSPVSR-----PESEAKDKKSP 1426
QY 56 KRSGE-----ELAGRMGKD---ANGH-----CLREGKEQ-----EGVK-- 87
DB 1427 FASGEASRPESVAESVDEACKAESRESIAKTHKDESSLDKAKEQESRRRESIAESIKPE 1486
QY 88 -----MATEGLARERRLORATKQKQLEEDGIPVPSVSRFTAPYHH----- 132
DB 1487 SGIDEKSALASKEASRPESVTDKSKPSRESIAESIAEKSTDEKSGAPPSKEASRPGSV 1546
QY 133 -KSTNAEEREAKEP-----VLKTHSKDVE-REAE--ICVDVDVVKMEPATNIIPEDAQD 183
DB 1547 VESVKDETEKSKPSRESIAESAKPPIEFREVSRPESVIDGIKDESA-----KPESRRD 1601
QY 184 A--AERGAARPPAVNSSYLPPLPWKGRGLGYACLNITYLNKAK-PPIFSRTCRMASIV----- 236
DB 1602 SPLASKEASRPESV-----LESVKDEPIKSTKSRRESVAESFK 1640
QY 237 -----DHRHPLQFED--EPEHHLKNPD---KSKPEQDELGHKFVQEL----- 274
DB 1641 ADSTKDEKSLTSDISRPEASAVENVDAPFKETSRPESAVGSKMDSKPSRRRESVK 1700
QY 275 -GLANARDIVK--MLCWNEYKIGIRFLR-----LSSEMFPPFASHPVHGK 315
DB 1701 DGAQSRSTSPASVAESAKDGDADDLKELSRPESTTOSKEAGSIKDEKSPLEASE--EASR 1758
QY 316 LAPFASEVLAEAGVAALGHLRHTTHFGQFTQLGSPKVEVESAIIRDLEYHDELLSLKL 375
DB 1759 PASVAESVKDAEAKSKESRRRESVAESKSPSPKSPASPAVAESIKD-----EAEKS 1810
QY 376 PEOONDAVMIIHMGOGFGDKAATLRFKRNRYARLSQCKNRLVLENDYDGVTVHOLLPV 435
DB 1811 KEESRRSV-----AEKSPSPKSPASPAVAESIKDEAKSKEE-----SRRESV 1856
QY 436 CEELNIPWLDYHHNICFPAHLREGTLDISDPLQERIANTWKRGIKQKMHYSEPCD 495
DB 1857 AEKSPPL-----SKASRSPASVAESIKDEAKSKEESRRRESVAESKSPSPKSPAS 1909
QY 496 GAVTFRD-----RRKHRPRVMTLPFCPP-----DMDLMTAEAKDKEQAFELMRTFKLPG 544
```

```
DB 1910 VAESIKDEAKSKESRRRESVAESKSPSPKSPASPAVAESIKDEAKSKESRR----- 1964
QY 545 FEKINDMVIYDRDENRKPAPPV-----KAPKKKGKGRKRTTD-----EAAAPPEEYDT 593
DB 1965 -ESVAESKSPSPKSPASPAVAESIKDEAKSKESRRRESVAESKSPSPKSPASPAV-- 2021
QY 594 AADDVKDAPEGPKVEPEERAMGPGPNRYWPLGCGEEMWLKPKKREVKKGKVPVEVEDEGE 653
DB 2022 -AESIKDEAKSKESRRRESV-----AEKSPSPKSPASPAVAESIKDEAE 2067

RESULT 11
Q9JXX5 ID Q9JXX5 PRELIMINARY; PRT: 1763 AA.
AC Q9JXX5;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE HISTONE ACETYLTTRANSFERASE QUERKOPF.
GN MORE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
KC TISSUE=BRAIN;
RA Thomas T., Voss A.K., Chowdhury K., Gruss P.;
RT "Querkopf, a MYST family histone acetyltransferase, is required for
normal cerebral cortex development.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF222800; AAF26744.1; -.
DR MGI:1858746; Morf.
DR InterPro: IPR001386; Linker_histone.
DR InterPro: IPR002717; MOZ_SAS.
DR InterPro: IPR001965; PHD.
DR Pfam: PF01853; MOZ_SAS; 1.
DR Pfam: PF00628; PHD; 2.
DR SMART: SM00526; H15; 1.
DR SMART: SM00249; PHD; 2.
DR Transferase.
KW TRANSFERASE.
SQ SEQUENCE 1763 AA; 196879 MW; 3A570C624B8094B0 CRC64;

Query Match 4.4%; Score 152.5; DB 11; Length 1763;
Best Local Similarity 20.0%; Pred. No. 0.1;
Matches 153; Conservative 86; Mismatches 232; Indels 293; Gaps 40;

QY 80 KQDEGVKMAIEGLARMERRLQRTKROKKQLEEDGIPVPS---VVSRRPTAPYHHKSTN 136
DB 319 RPKKGRKLLHEKAAQIKRKYAKPIGRPKNLKQRLSVTSDEGMSAFTGRGSPDTDIK 378
QY 137 AEERAKEFVLKTHSKDVEREAEIGVDVVKM-EPAATNILEPDAQDAEAERGAARPPAV 195
DB 379 ISIKOESADVSLVGNKLTVEEDL---DVFOAQELSEKTECESGVE---DCGRYFSV 431
QY 196 -----NSYPLPLPWKGRIGYACL-----NTYLRNAK-----PP----- 223
DB 432 IEFGKYEIQTWYSSPYQPYEARLP-KLYLCEFCLYKMKSNILLRHKKCGWGFHPANEI 490
QY 224 -----IFSSRTCRMASI-VDIR-----HPLQF-----EDEPHHLKN 254
DB 491 YRRKDLVSFEVDGNNSKIYQCNLCLLAKLFLDHKTLTYDYDVEFPFLYVLTKNDEKCHLVG 550
QY 255 KPDKSKPEQDELGHKFVQELGLANARDIVKMLCWNEYKIGIRFLRSLSEMFPPFASHPVHG 314
DB 551 YFSKEK-----LC-QQKYNVSCIMLMPQ-----HQROGF 578
QY 315 KLAPFASEVLAEAGVAALGHLRHTTHFGQFTQLGSPKVEVE-----SAIRDL 363
DB 579 -----GRFLIDFSYLLSKREG---QAGSPEKPLSLDGLRSLYLAWKSVI--L 620
```


George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Abayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler C., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahle K., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Laško P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong R.N., Zhou X., Zhu S., Zhu H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of *Drosophila melanogaster*."
 RT Science 287:2185-2195(2000).
 RL EMBL: AE003713; AAF55319.1;
 DR FlyBase: FBgn0038433; CG14896.
 SQ SEQUENCE 828 AA; 91444 MW; 96B56F321CB2DC6 CRC64;

Query Match 4.1%; Score 143.5; DB 5; Length 828;
 Best Local Similarity 19.9%; Pred. No. 0.16;
 Matches 162; Conservative 99; Mismatches 279; Indels 273; Gaps 40;
 QY 2 PRRKAAALDTPQSESTFSTLSDSSAPARNLRSGRNILQPSSEKDR-DHEKRSRG 59
 DB 77 PRRKAAAKETNTSDKKEKTDVILAPVKE-----VPSLSIKNSDHSKSG 126
 QY 60 -----ELAGRMCKDANGCHLREGKEQ-----EGVKMAIEGIARM----- 96
 DB 127 POEPAPVATLNISKEQDPNDVNPVNSSENMAPKPKQERPKRVRKASEDIVEVNAID 186
 QY 97 -----ERLQRA-----TKRQKQLE-----EDGIPVPSVSRPPTA 128
 DB 187 AINPTALASVOEDHLNQAODEMPKPKRKSALMLANVEEHAKEDHLERKRGKPSL 246
 QY 129 PYHKTNAER-----EAKPEVL-----KTHSKOVE- 155
 DB 247 DTDHVSQDVPKPKRGTTPADRETQPLGDEKQEPSLPVVEPAKTRRNARKASAEV 306
 QY 156 -----RAEIGVDVVKMEPAANTILPEDAQAAEAGAARPAPVNSSYLPPLPWKG 206
 DB 307 VSTSGEHLAQIATEPTEVTESEPAQSELL-PEEGEENKTRRRGRKPTVDTEGP 362
 QY 207 RLGYACLTNLYRNAP-PFESSRTRMASIVDHRHPLQFDE-----PEHLKNKP 256
 DB 363 -----KTPAEFVPLPSHRRGRKAT-----EDEVLPADLAVPTKURGR- 404
 QY 257 DKSKEPDELGHKTVQELGLANARDIVKMLCWNKYGIRFLRSLSEM---FPFASHPV-H 312
 DB 405 RASMEPE---HK---DELTASSEVVVELP---TAKTTRRGRKTSVDMEATPEKPKPSRR 456

QY 313 GYKLAFSEAEVLAAGVAAELGHLRLTHPCOF-----TOLGSPRKEVVEVSARD 362
 DB 457 GRKASASVDEQPAAKKTATRRGRKNEAHEDEERGHIDLQDLPTETASPLVDTSGSPSKA 516
 QY 363 LEYHDELLSLIK-----LPEQONKDAV--MIHMGQGFCDKAATLRFKRNRYARLSQCKN 416
 DB 517 SD-AEELTERRRGRNLPKRYEEAPDDEKPHSG-----LRRARKPAATKSLASK- 565
 QY 417 RLVLNDDVGTVHDLPLVCEELNIPMLVDYHHHNICFDPAAHLREGTLDISDPKLOERIA 476
 DB 566 --AEPDPVTPSPVNOPPVKSEDTP-----DNTVLSLEPTTSQR- 603
 QY 477 NTWKRGKIK-QKMHYSPPCDGAV--TPDRRKHR-----PRVMT----- 512
 DB 604 ----REGNLPKRYNTEDLDEMTTPARSRRVRLTAKALELIVDSSPRVTPKPKGKA 659
 QY 513 ----LPPC-----PDMOLMIKAKEQAVFELMRTFKLPGEKINDMVPYDRDD----- 558
 DB 660 ANSEEPKAKKPTPEFPSTTEAAGPEH---EPIPATKGRGTRRKADDTDLKPKDVKTAKT 716
 QY 559 -----ENRPAPPVKAPKKKKGKRRKTDEEAPEE-----VDTAADVDK 600
 DB 717 VRGAGGKTKVETETEKQPIKPP---RGARTKTPSEEAPOEEQVKKSAARSRAKGTGA 773
 QY 601 A-PEGKPEVPEERAM-----GGPYNRVYW 624
 DB 774 VEPEPAEDPQVEASFSTASVRRGGRARKVHF 806
 RESULT 14
 Q9H945
 ID Q9H945 PRELIMINARY; PRT; 914 AA.
 AC Q9H945;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE CDNA FLJ13022 FIS, CLONE NT2RP3000753, WEAKLY SIMILAR TO NEUROFILAMENT
 DE TRIPLET H PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Isodai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
 RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,
 RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,
 RA Masuho Y., Kanehori K.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
 DR EMBL: AK023084; BAB14396.1; --
 SQ SEQUENCE 914 AA; 100583 MW; 7E7A9B3A5BAA53C2 CRC64;
 Query Match 4.1%; Score 143.5; DB 4; Length 914;
 Best Local Similarity 19.3%; Pred. No. 0.18;
 Matches 156; Conservative 103; Mismatches 281; Indels 269; Gaps 39;
 QY 11 LDTPOSESSTF-----SSTLSDSSAPARNLRSGRNILQPSSEKDR-DHEKRSGE-IA 63
 DB 100 VDLPLLESSTFTDENPSETESEAAAGIGKLEGEDGV-KCLSEKDTYDTSIDSLLENLD 158
 QY 64 GRMMGK---DANCHCLREGKEQEGVKMAI-----EGLARMERLQR--ATKROKQLE 112
 DB 159 KKKKGGKFPFESDRCLRSOLSDSSSADRCNRQSSSACLEIKVPKNPSAKRSKKEGH 218
 QY 113 EDGIPVPSVSRPTAPYHHKSTNAEEREAKPEVLKTHSKDVEREAETGVDDVVKMEPAA 172
 DB 219 PGSTTPKGLL-----PDSFHTLETDEKPSVNERPSEKDAEQEGG---GIITRQL 269
 QY 173 TNLIFPEDAQDAEAGAARPPAVNSYLPPLPWKGLRGVACLTNLYRNAPKPPFSSRTCRM 232

```

Db 270 KNLKDEKVLGRGIFPSRDP-ITTAGOPLGERLEIIVQSKMDEKNAHIP-----S 320
QY 233 ASIVDRHPLQEDPEH-----HLKKNPKSKPEQDELGHKFFQELGL 276
Db 321 ESTACKRDEQAKEPGHIPQHVEAVNEVDNTOOKDDSDAPCSSLG-----LSS 374
QY 277 ANARDIVKMLCNWRYGTRFLRSLSEMPFFASHPVHGYKKLAPP-ASEVLAAGRAVAAELG 335
Db 375 SSGGDAARA-----PKSVPRKRLTSTYNLR-----HAHSLGLDASKVTSE--KEAAQVN 424
QY 336 HRLTHPGQFTOLGSPRKE-----VYESAIRDLEYHDELLSLKLLPEOONRDVAMIHM 390
Db 425 PIMPRENGA-SESGDPLDEDDVTVVDEQPKFMEW-----CAEENQOE--LIANFN 472
QY 391 GQP-----GDKAATLFRKRYARLSOSCKNRLVLENDVDVGW 427
Db 473 AQYMKVQKGIQLEKEGQPTPRARKSDKLKEIWKSKKR-----SRKCRSSLESOK----- 523
QY 428 TVHDLPLVCEELNIPWLDYHHNIC-----CSPQMLFMTWFKLSNCKWFFLETTETRSIVIVKKLNTRLPGGVPPVKHPLQ 575
Db 524 -----KSPQMLFMTWFKLSNCKWFFLETTETRSIVIVKKLNTRLPGGVPPVKHPLQ 575
QY 454 -FDPAHLREGTLDISPKLQERI-----ANTWKRKGKQKMHYS-----EPCDGA-V 498
Db 576 KYAPSSLYPSSLOAE--RLKKHLKFFPGATPAKNWKKMKLWAKFRENPDQVPEPDSGV 633
QY 499 TP-----RDRKRHRPRVMTLP-----PCPP 518
Db 634 SPGNSEDSTEEVKEDKDRNSHPANLPTPASTRILRKYSNIRGKLRAQQLIKNEKMECPD 693
QY 519 DMDLMEAKDKQAVFELMRTKLPFGKINDM-----INLMSPKLALQVDAGFPVKPKSTEGMKGR 738
Db 694 --ALAVESKPSRKSVC-----INLMSPKLALQVDAGFPVKPKSTEGMKGR 738
QY 569 -----PKKKGGKRRRTTDEAAEPEVDTAADVDKDAPE-----GPKVEPEERAM 615
Db 739 KGQVSEILPKAEVQSKRRT---EGSSPDCKNKGPTVKASKEKHADGATKTPAAKRP- 794
QY 616 GGPYNRVYPLGCEEWLKPCKREVKGKV 644
Db 795 -AADRSSQP-----PKKTSLEKNKY 814

RESULT 15
Q91255 PRELIMINARY; PRT: 1110 AA.
AC Q91255;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NF-180.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyporoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-CENTRAL NERVOUS SYSTEM;
RX MEDLINE=95287814; PubMed=7770000;
RA Jacobs A.J., Kamholz J., Seizer M.E.;
RT "The single lamprey neurofilament subunit (NF-180) lacks
RT multiphosphorylation repeats and is expressed selectively in
RT projection neurons."
RL Brain Res. Mol. Brain Res. 29:43-52(1995).
DR EMBL; U19361; AAA80106.1;
DR InterPro; IPR000533; Tropomyosin.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament. 1.
DR PRINTS; PR00194; TROPOMYOSIN.
DR PROSITE; PS00226; IF; UNKNOWN.1.
SO SEQUENCE 1110 AA; 123817 MW; 6558DA73DAF6974C CRC64;

```

```

Query Match 4.1%; Score 141.5; DB 13; Length 1110;
Best Local Similarity 17.6%; Pred. No. 0.31;
Matches 132; Conservative 102; Mismatches 275; Indels 239; Gaps 24;

QY 2 PSRKSAAALDTP-----QSESSTFSTLDSSAPSPARNLRRSGRNILQPSSEKDRD 53
Db 429 PSAPKAAKVEKKVSKKPEIKVESEPISAOLDTDLDLAOEVEYMEAKAAPVVSAAKDEE 488
QY 54 HE-----KRSGEELAGRMKGDKANGHCLREGKFOEGVKMATEGLRMRRL 100
Db 489 EEEEEEEKEEEAEAEEDRGRKEGAEEAEAEVEEKEEAEAEAEAEAEAEAEAEAE 548
QY 101 QRATKROK-----KOLEDDGI--PVPVSYSRPTAPYHHKSNABEAE 142
Db 549 EAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 608
QY 143 KEPVLKTHSKDVERAEALGVDDVVKMEPATINIIPEDAODAAERGAARPPAVNSSLPL 202
Db 609 EEEVEAEATKE-EVEAEAEVEE-----EGEAEAEAEAEAEAEAEAEAEAE 651
QY 203 PWKGRGLGVACLTLYLRNAKPPIFSRTCRMASIVDHRHPLQFDEPEHHLKKNPKDSKE- 261
Db 652 -----KAKTQAEVEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 700
QY 262 --PODELGHKFFVQELGLANARDIVKMLCNWKEYGIRFLRSLSEMPFASHPVHGYKLAPP 319
Db 701 DAEDEAEAEVEEKE-----EETVK-----SD 721
QY 320 ASEVLAAGRAVAAELGHLRTHPGQFTOLGSPRKEVVESEAIRDLEYHDELLSLKLP 379
Db 722 AEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 769
QY 380 NRDAVMIHMGQFGDKAATLERKRNRYARLSQCKNR-----LVLENDVDVGW 427
Db 770 AKEA-----SDDEKPEEEVKESEAPVPAEAKKAPKAPKAPKAPKAPKAPK 814
QY 428 TVHDLPLVCEELNIPM--VLDYHHHNICFDPALHREGTLDISDPKLQRIANTWKRGIK 485
Db 815 -----PTSEPEDEPKAEVVE-----KKGAEAPKPKAPKAPKAPKAPKAPK 854
QY 486 QKMHYSE-----PCDGAVTPTDRRRKRP-----RVMTLPCCPPMDLMI 524
Db 855 KEEPEEPEEPTEEEPKPAAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK 914
QY 525 EAKDKEQAVFELMRTKLPFGKINDMVPYD-----RDDENRPPAPVVKAPKKK 573
Db 915 EKEDEDEEE-----EEVEVKPEDAKPVKSKPAPAKAEDEDEPKAPKAPKAPK 962
QY 574 GGRKRRTTDEAAEP-----EEVDTAADVDKADAPGPKPEVPEEERAMGGPYNRVWP 625
Db 963 KPARPKEEPEDKAEPAKKEKHSPEERKPIKETAKPAKAPKAPKAPKAPKAPKAPK 1011
QY 626 LQCEEWLKPCKREVKGKVPEEVEDEGE 653
Db 1012 -----AEPKKIEVKVKVTKKVVVEIE 1033

Search completed: January 15, 2002, 14:06:25
Job time: 964 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:52:22 ; Search time 74.94 Seconds
(without alignments)
377.112 Million cell updates/sec

Title: US-09-724-296-4
Perfect score: 1979
Sequence: 1 DDHAPREMFDCDKRPIPWG.....KRLTARKRSRKEVEDEK 371

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1979	100.0	599	S71134	UV-endonuclease -
2	960	48.5	656	S55262	UV-endonuclease -
3	360	18.2	320	S55418	UV-endonuclease ho
4	306.5	15.5	322	E83974	hypothetical prote
5	229.5	11.6	326	C75350	probable UV damage
6	111.5	5.6	517	S45996	H+-transporting AT
7	110.5	5.6	415	B42254	H+-transporting AT
8	110	5.6	465	T44675	H+-transporting AT
9	108.5	5.5	492	A31886	H+-transporting AT
10	108	5.5	500	T14363	probable H+-transp
11	104.5	5.3	472	F69395	H+-transporting AT
12	104	5.3	463	T44310	H+-transporting AT
13	104	5.3	490	S35167	H+-transporting AT
14	103	5.2	494	S48395	H+-transporting AT
15	103	5.2	494	S24387	H+-transporting AT
16	102	5.2	513	C44138	vacuolar H+-transp
17	101.5	5.1	465	C75028	H+-transporting AT
18	101.5	5.1	503	JC4198	adenosinetriphosph
19	101.5	5.1	588	JC7206	phosphoprotein pho
20	101	5.1	491	T34226	hypothetical prote
21	101	5.1	800	B59106	hypothetical prote
22	101	5.1	1145	T33606	hypothetical prote
23	100.5	5.1	511	B44138	hypothetical prote
24	100	5.1	373	C70355	H+-transporting AT
25	100	5.1	513	G70355	8-amino-7-oxonon
26	99.5	5.0	273	A33281	H+-transporting AT
27	99.5	5.0	635	E64336	hypothetical prote
28	98.5	5.0	507	T06619	hypothetical prote
29	97.5	4.9	487	T05679	H+-transporting AT

97.5 4.9 511 2 S32614 H+-transporting AT
97.5 4.9 520 2 G96788 protein T4012.24
32 97 485 2 C86336 hypothetical prote
33 96.5 4.9 465 2 F71213 probable H(+)-tran
34 96.5 4.9 1181 2 T01799 hypothetical prote
35 96 4.9 337 2 S74850 hypothetical prote
36 96 4.9 501 2 S75951 hypothetical prote
37 96 4.9 511 2 S13080 (hypothetical prote
38 94.5 4.8 523 2 D86492 H+-transporting AT
39 94.5 4.8 523 2 D72130 hypothetical prote
40 94 4.7 512 2 T01861 hypothetical prote
41 94 4.7 519 2 D81293 hypothetical prote
42 93.5 4.7 648 2 S16866 probable type II p
43 93.5 4.7 665 2 E70118 gene H19 protein -
44 93 4.7 462 2 G70015 flagellar hook-ass
45 93 4.7 465 2 F69227 conserved hypothet
ATP synthase, subu

AKAO

TRV:ICATFD

ALIGNMENTS

RESULT 1
S71134
UV-endonuclease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 20-Jun-2000
C:Accession: S71134.139815
R:Takao, M.; Yonemasu, R.; Yamamoto, K.; Yasui, A.
Nucleic Acids Res. 24, 1267-1271, 1996
A>Title: Characterization of a UV endonuclease gene from the fission yeast Schizosacch
A:Reference number: S71134; MUID:96188860
A:Accession: S71134
A>Status: Preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-599 <TAK>
A:Cross-references: EMBL:D78571; NID:g1399000; PIDN:BAAL1415.1; PID:g1399001
A:Experimental source: strain Sp972
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Lucas, M.; Gaillardin, C.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z21881
A:Accession: T39815
A>Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-599 <LYN>
A:Cross-references: EMBL:AL023859; PIDN:CAA19577.1; GSPDB:GN00067; SPDB:SPBC19C7.09c
A:Experimental source: strain 972h-; cosmid c19C7
C:Genetics:
A:Gene: UVDE; SPBC19C7.09c
A:Map position: 2

Query Match	Best Local Similarity	Score	DB 2	Length
Matches 371; Conservative	100.0%;	100.0%;	0; Indels	0; Gaps
0; Mismatches	0; Indels	0; Gaps	0; Indels	0; Gaps
QY 1	DDHAPREMFDCDKRPIPWGRLGVACLTILRSKRVFCSRTCRITTIORDGLSVKQL	60		
Db 229	DDHAPREMFDCDKRPIPWGRLGVACLTILRSKRVFCSRTCRITTIORDGLSVKQL	288		
QY 61	GTQNVLDLILKLVENHNFGTHFMRVSSDLFPFAHQSHLEEVGKLANKYN	120		
Db 289	GTQNVLDLILKLVENHNFGTHFMRVSSDLFPFAHQSHLEEVGKLANKYN	348		
QY 121	HLTMTHPGQYQTASPREVVDSAIRDLAYHDEILSRMKLNEQLNKDAVLIHLGGTFEG	180		
Db 349	HLTMTHPGQYQTASPREVVDSAIRDLAYHDEILSRMKLNEQLNKDAVLIHLGGTFEG	408		
QY 181	KKETLDRFRKNYQRLSDSVKARLVLENDVSVSWQDILLPLCQELNIPVLWDHNNHNVPC	240		
Db 409	KKETLDRFRKNYQRLSDSVKARLVLENDVSVSWQDILLPLCQELNIPVLWDHNNHNVPC	468		
QY 241	TLREGSLDLMLPITPTRETWRKGTQKHYSADPTAISGMKRAHSDRVDFPPCPD	300		

Db 469 TLREGSLDMLPIPTIRETWRTKGTQKHYSADPTAISGMKRRASDRVDFPPCDP 528
 Qy 301 TMDLMEAKEQAVFELCRVYELQNPCCPLEIMGPVDTQDCGYPPGAEKRLTARKRR 360
 Db 529 TMDLMEAKEQAVFELCRVYELQNPCCPLEIMGPVDTQDCGYPPGAEKRLTARKRR 588
 Qy 361 SRKEVEDEK 371
 Db 589 SRKEVEDEK 599
 RESULT 2
 UV-endonuclease - Neurospora crassa
 C:Species: Neurospora crassa
 C:Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
 C:Accession: S55262
 R:Yajima, H.; Takao, M.; Yasuhira, S.; Zhao, J.H.; Ishii, C.; Inoue, H.; Yasui, A.
 EMBO J. 14, 2393-2399, 1995
 A:Title: A eukaryotic gene encoding an endonuclease that specifically repairs DNA damage
 A:Reference number: S55262; MUID:95292980
 A:Accession: S55262
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-656 <YAN>
 A:Cross-references: EMBL:D11392; NID:gl526560; PID:g927215
 C:Genetics:
 A:Gene: uvel

Query Match 48.5%; Score 960; DB 2; Length 656;
 Best Local Similarity 45.4%; Pred. No. 8.2e-67;
 Matches 205; Conservative 54; Mismatches 95; Indels 98; Gaps 10;
 Qy 15 PIPWRGLGYACLTILSRMKRVFCSTCRITTI----- 49
 Db 201 PLPWKRLGYACLTILSRNKPPIFSSRTCRMASIVDRHRLQFDEPEHLLKPKDSK 260
 Qy 50 ---QRGLSVKQLGQNVLDLILKLVWNNHFGIHFMRVSSDLFPFASHAKYGYTL-BFA 105
 Db 261 EPODELGHKRVQELGANARDIVKMLCWEKYGIRFLRLSSSEFFAFSHVPGYKLAPPA 320
 Qy 106 QSHLEEVGKLANKYHRLTHHPGOYTQIASPREVVVDSAIRDLAYHDEILSRMKLNQIN 165
 Db 321 SEVLAEAGRAVAELGHLRHTHPGOYFQLGSPKRVVSAIRDLVHDELLSLKLPQON 380
 Qy 166 KDAVLIHLGGTTEGKETLDRFRKNYQRLSDSKARLVLENDVSVQDILLPLCOELN 225
 Db 381 RDVAVIHHGGQFGDKAATLEREFKNYARLSOSCKNRLVLENDVGVTHVHLLPVCEELN 440
 Qy 226 IPLVLDWHHHNIV--PQLREGSLDLM--PLIPTIRETWRTKGTQKHYSADPTAIS 281
 Db 441 IPWLDYHHNICFDPAHLREGTLDISDPKQERIANTRKKGIGKMHYSEPCD-GAVT 499
 Qy 282 GMRKRAHSDRVDFPPCDPMDLMEAKEQAVFELCRVYEL-----Q 325
 Db 500 PRHRRKRRPRVMTLPPCPDMDLMEAKEQAVFELCRVYEL----- 559
 Qy 326 NPCCP-----LEIMGPE-----YDQTRDG----- 344
 Db 560 NRPAPPVAPKPKKKKKRKTITDEAAEPPEVDTAADDVYKDAPEGKVPPEERAMGGPY 619
 Qy 345 ---YPPCAEKLRTARKRRSRK-----EEVED 369
 Db 620 NRVMYPLGCEEWLKKPKKRVKKGKVPPEVEDE 651
 RESULT 3
 UV-endonuclease homolog ywjd - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 15-Oct-1999
 C:Accession: S55418; B70060

R:Glaser, P.; Danchin, A.
 submitted to the EMBL Data Library, May 1995
 A:Description: Cloning and sequencing of the Bacillus subtilis chromosomal region fro
 A:Reference number: S55414
 A:Accession: S55418
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-320 <GLA>
 A:Cross-references: EMBL:249782; NID:g853752; PID:CAA9865.1; PID:g853757
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C:Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.: Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
 lect, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, J.; Sato, T.; Scanl
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, Y.; Sekowska, A.; Se
 keuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram positive bacterium Bacillus subtili
 A:Reference number: A69580; MUID:98044033
 A:Accession: B70060
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-320 <KUN>
 A:Cross-references: GB:299123; GB:AL009126; NID:g2636240; PID:NCBI5748.1; PID:ell1862
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: ywjd

Query Match 18.2%; Score 360; DB 2; Length 320;
 Best Local Similarity 29.8%; Pred. No. 1.5e-20;
 Matches 84; Conservative 61; Mismatches 109; Indels 28; Gaps 8;
 Qy 48 TTORDGLSVKQLGQNVLDLILKLVWNNHFGIHFMRVSSDLFPFASH--AKYGYTLEFA 105
 Db 34 TERKEALLTVTKANLRNTM---RTLHVLIHGILPYRFSSSIVPLATHPDMVMDVTFPF- 89
 Qy 106 QSHLEEVGKLANKYHRLTHHPGOYTQIASPREVVVDSAIRDLAYHDEILSRMKLNQIN 165
 Db 90 QKEFEIGELVKTHQLTSPFHPNQFTLTSPKESVTNAVTDMAHYRMLEAMGIADR-- 147
 Qy 166 KDAVLIHLGGTTEGKETLDRFRKNYQRLSDSKARLVLENDVSVQDILLPLCOELN 225
 Db 148 --SVINIHIGGAYGNKDTATATQAFHONIKOLPQEKERMTELENDKTVTTEETLQVCEQED 205
 Qy 226 IPLVLDWHHHNIVPGTTLREGSLDLMPLIPTIRETWRTKGTQKHYSADPTAISGMKR 285
 Db 206 VPFVDFDHFHFVNP----DDHADLNVALPRMIKTWERIGLOPKVHLSPKSEQAI----- 256
 Qy 286 RAHSDRV---PDFPPCD-----PTMDLMEAKEQAVFEL 318
 Db 257 KSHADYVDANFLLPILLERFROWGTNIDEMIEAKOKDKALLRL 298

RESULT 4
 E83974
 hypothetical protein BH2597 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
 C:Accession: E83974
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
 A:Reference number: AB3650; MUID:20263314
 A:Accession: E83974
 A:Status: preliminary
 A:Molecule type: DNA

[illegible]

A:Description: Analysis of the gene encoding the 57 KD B-subunit of the vacuolar H⁺-ATPase
A:Reference number: S25167
A:Accession: S25167
A:Molecule type: mRNA
A:Residues: 1-490 <DOW>
A:Cross-references: EMBL:X67839; NID:g6809; PIDN:CAA48034.1; PID:g8810
C:Genetics:
A:Gene: FlyBase:Vha55
A:Cross-references: FlyBase:FBgn0005671
C:Superfamily: H⁺-transporting ATP synthase alpha chain; H⁺-transporting ATP synthase
C:Keywords: ATP biosynthesis; hydrolase
F:210-381/Domain: H⁺-transporting ATP synthase alpha chain homology <ATP>

Query Match 5.3%; Score 104; DB 2; Length 490;
Best Local Similarity 19.7%; Pred. No. 2.4; Indels 86; Gaps 13;
Matches 61; Conservative 48; Mismatches 114

QY 7 EMFDCLDKPI-TW-----RGRUGYACLNLTILRSMKRVF-----CSRTCR 45
DB 128 DFLDIQOQPINWRIYPEEMTGTGSAIDVMSIARGOKIPFSAAGLPHNEIAAQICR 187
QY 46 ITTITORDGLSVKQGTQNVLDLIKLVENHNHFGIHFMRVSSDLFFPASHAKYGYTLEFA 105
DB 188 -----QAGL--VKLPG-KSVLD-----DHTDNEFAIVFAAGVNM-----ETAREF 224
QY 106 QSHLEEVGKLANK-----YNHRLTMHPGQYQTASPREVV-----D 142
DB 225 KQDFEENGSMENVCLEFLNLANPTIERTIPLRLATAEFLAYOCEKHVLVILTMSSVA 284
QY 143 SATRDL-AYHDEILSRMKLEQNLKDAVLIIHLGGTFECKKTELDRFRKNYORLSDSVKA 201
DB 285 EALREVSAAAREEVPGRGPGYMYTDLATYERAGRVEGRNGSITO-----PVNVLPSRLMKSAI 386
QY 202 RLVLENDVSWSDVODLLPLCOELNIPVLVDWHHHNIVPGTLREGSLDLMPLIPTIRETWT 261
DB 333 ILTMPDDDTHPIPDLTGYITEGOIYVDRQLHNRQIYP-----P 332
QY 262 RKGITQKQH 270
DB 387 GEGMTRKDH 395

RESULT 14
S18395
H⁺-transporting ATPase (EC 3.6.1.35) chain B, vacuolar - tobacco budworm
C:Species: Heliothis virescens (tobacco budworm)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jan-2000
C:Accession: S18395
R:Gill, S.S.; Ross, L.S.
Arch. Biochem. Biophys. 291, 92-99, 1991
A:Title: Molecular cloning and characterization of the B subunit of a vacuolar H⁺-ATPase
A:Reference number: S18395; MUID:92027804
A:Accession: S18395
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-494 <GIL>
A:Cross-references: GB:S61797; NID:g237458; PIDN:AAB20098.1; PID:g237459
C:Superfamily: H⁺-transporting ATP synthase alpha chain; H⁺-transporting ATP synthase
C:Keywords: hydrolase
F:214-385/Domain: H⁺-transporting ATP synthase alpha chain homology <ATP>

Query Match 5.2%; Score 103; DB 2; Length 494;
Best Local Similarity 19.7%; Pred. No. 2.9; Indels 86; Gaps 13;
Matches 61; Conservative 48; Mismatches 114

QY 7 EMFDCLDKPI-TW-----RGRUGYACLNLTILRSMKRVF-----CSRTCR 45
DB 132 DFLDIQOQPINWRIYPEEMTGTGSAIDVMSIARGOKIPFSAAGLPHNEIAAQICR 191
QY 46 ITTITORDGLSVKQGTQNVLDLIKLVENHNHFGIHFMRVSSDLFFPASHAKYGYTLEFA 105

A:Description: Analysis of the gene encoding the 57 KD B-subunit of the vacuolar H⁺-ATPase
A:Reference number: S25167
A:Accession: S25167
A:Molecule type: mRNA
A:Residues: 1-490 <DOW>
A:Cross-references: EMBL:X67839; NID:g6809; PIDN:CAA48034.1; PID:g8810
C:Genetics:
A:Gene: FlyBase:Vha55
A:Cross-references: FlyBase:FBgn0005671
C:Superfamily: H⁺-transporting ATP synthase alpha chain; H⁺-transporting ATP synthase
C:Keywords: ATP biosynthesis; hydrolase
F:210-381/Domain: H⁺-transporting ATP synthase alpha chain homology <ATP>

Query Match 5.3%; Score 104; DB 2; Length 463;
Best Local Similarity 18.9%; Pred. No. 2.2; Indels 92; Gaps 13;
Matches 70; Conservative 55; Mismatches 154

QY 27 LNTILSRMKRVFCSRTCRITITORDGLSVKQGTQNVLDLIKLVENHNHFGIHFMRVS' 86
DB 139 MNTLVRGOKLPFGSGGLPHNMIAQIAQAKVLGEE-----EQFAVVF----- 182
QY 87 SOLFPFASHAKYGYTLE---FAQSHLEEVGKLANK-----YNHRLTMHPG 128
DB 183 -----AAMGITPEAFNFKFSFEETGAIERAVLFNLADDPATERTIIPRMALTV 233
QY 129 QYQTASPREVV-----DSAIRDL-AYHDEILSRMKLEQNLKDAVLIIHLGGTFE 179
DB 234 EYLAFLYDMQVLVILTMNTYAEALREISAAREEVPGRGPGYMYTDLATYERAGRVR 293
QY 180 GKKETLDRFRKNYORLSDSVKARLVLENDVSWSDVODLLPLCOELNIPVLVDWHHHNIVP 239
DB 294 GKKGSIQW-----PILTMPDDDTHPIPDLTGYITEGOIYVLSRELHRKGIYP 341
QY 240 GTLREGSLDLMPLIPTIRETWTQKQKOH-----YSES---ADPTAISGMKRR 286
DB 342 -----PIDVLSRLMKMDGKGTREDHPQLSQOLYAAYAEGRSLRDLVAVVGEAL 395
QY 287 AHSRDFVDFPCDPTMDLMIAEKEQAVFE-LCRRYELQNPCCPLETNGPEYDQTRDGY 345
DB 396 SETDRYKLFADRFERFEFAQRYEDRSIFETLDLGWELL-----AELPESELKRVKREY 450
QY 346 ---YPPGAERK 353
DB 451 ILKYHPKRRK 461

RESULT 13
S25167
H⁺-transporting ATPase (EC 3.6.1.35) chain B, vacuolar - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C:Accession: S25167
R:Dow, J.A.T.; Kaiser, K.
submitted to the EMBL Data Library, July 1992

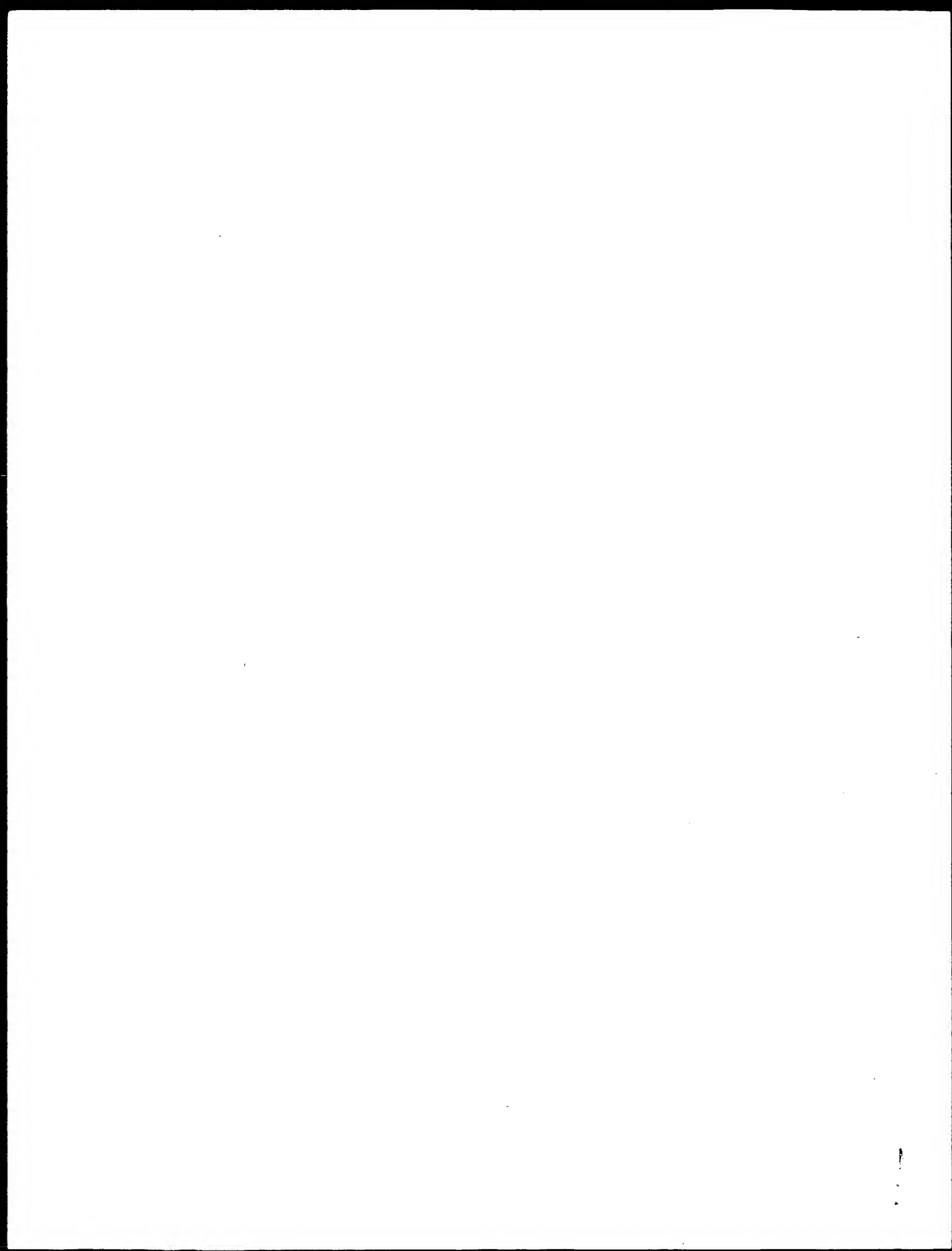
```
Db 192 -----QAGL--VKIPG-KSVLD-----DHEDNFAIVFAAMGVNM-----ETARFF 228
Qy 106 OSHLEEVGKLANK-----YNHRLTMHPGQYQTASPREVV-----D 142
Db 229 KODFEENGSMENVCLFLNLANDPTIERIITPRALTAAEFLAYQCEKHVLVILTMSSYA 288
Qy 143 SAIRD-LAYHDEILSRMKLNQKDAVLIHLGCTEGCKETLDRFRKNYQRLSDSVKA 201
Db 289 EALREVSAAAREEVPGRGFGYMYTDLATYERAGRVEGRNGSITQI-----P 336
Qy 202 RLVLNDDVSWSDLLPLCOELNPLVLWDHWHHNIIVPGTLREGSLDMLPIPTIRETWT 261
Db 337 ILTPNDIDITHPIPDLTGYITEGQIYVDRQLHNRQIYP-----PNNVLPSSLRLMKSAI 390
Qy 262 RKGITQKQH 270
Db 391 GEGMTRKDH 399

RESULT 15
S24387
H+-transporting ATPase (EC 3.6.1.35) chain B, vacuolar - tobacco hornworm
C:Species: Manduca sexta (tobacco hornworm)
C:Date: 09-Jun-1994 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C:Accession: S24387; S19206
R:Novak, F.J.S.; Graef, R.; Waring, R.B.; Wolfersberger, M.G.; Wieczorek, H.; Harvey, W.
Biochim. Biophys. Acta 1132, 67-71, 1992
A:Title: Primary structure of V-ATPase subunit B from Manduca sexta midgut.
A:Reference number: S24387; MUID:92379093
A:Accession: S24387
A:Molecule type: mRNA
A:Residues: 1-494 <NOV>
A:Cross-references: EMBL:X64354; NID:g9713; PIDN:CAA45706.1; PID:g9714
C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase al
C:Keywords: ATP biosynthesis; hydrogen ion transport; hydrolase; membrane-associated com
F:214-385/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

Query Match 5.28; Score 103; DB 2; Length 494;
Best Local Similarity 19.7%; Pred. No. 2.9;
Matches 61; Conservative 48; Mismatches 114; Indels 86; Gaps 13;

Qy 7 EMFDCLDKPI-PW-----RRLGYACLNILRSMKERVF-----CSRTCR 45
Db 132 DFLDIQGPINPWSRIYPEEMIQGISADVMSIARGQKIPISAAGLPHNEIAAQICR 191
Qy 46 ITTIORDGLESVKQLGTQNVLDLIKLVENHNFHGMVSVSSDLFPFASHAKYGYTILEFA 105
Db 192 -----QAGL--VKIPG-KSVLD-----DHEDNFAIVFAAMGVNM-----ETARFF 228
Qy 106 OSHLEEVGKLANK-----YNHRLTMHPGQYQTASPREVV-----D 142
Db 229 KODFEENGSMENVCLFLNLANDPTIERIITPRALTAAEFLAYQCEKHVLVILTMSSYA 288
Qy 143 SAIRD-LAYHDEILSRMKLNQKDAVLIHLGCTEGCKETLDRFRKNYQRLSDSVKA 201
Db 289 EALREVSAAAREEVPGRGFGYMYTDLATYERAGRVEGRNGSITQI-----P 336
Qy 202 RLVLNDDVSWSDLLPLCOELNPLVLWDHWHHNIIVPGTLREGSLDMLPIPTIRETWT 261
Db 337 ILTPNDIDITHPIPDLTGYITEGQIYVDRQLHNRQIYP-----PNNVLPSSLRLMKSAI 390
Qy 262 RKGITQKQH 270
Db 391 GEGMTRKDH 399
```

Search completed: January 15, 2002, 13:52:23
Job time: 392 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:06:13 ; Search time 134.53 Seconds
(without alignments)
403.382 Million cell updates/sec

Title: US-09-724-296-4
Perfect score: 1979
Sequence: 1 DDHAPREMFCDLKPWRG.....KRLTARKRRSRKEEVEDEK 371

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1979	100.0	599	3 Q10988	Q10988 schizosacch
2	960	48.5	656	3 Q01408	Q01408 neurospora
3	359	18.1	317	2 Q914F0	Q914F0 bacillus ce
4	359	18.1	325	2 Q914E9	Q914E9 bacillus ce
5	306.5	15.5	322	2 Q9K9P8	Q9K9P8 bacillus ha
6	229.5	11.6	305	2 Q9S0M9	Q9S0M9 deinococcus
7	229.5	11.6	326	2 Q9RTE6	Q9RTE6 deinococcus
8	111	5.6	496	5 Q9XYC8	Q9XYC8 aedes aegypt
9	109	5.5	492	5 Q44427	Q44427 culicx quing
10	108	5.5	489	5 Q9U6A8	Q9U6A8 carcinus ma
11	105	5.3	509	5 Q96066	Q96066 ascidia syd
12	104	5.3	881	5 Q9GR11	Q9GR11 paramacium
13	103	5.2	502	13 Q9W6M4	Q9W6M4 oncorhynch
14	102.5	5.2	488	10 Q9ASE0	Q9ASE0 oryza sativ
15	101.5	5.1	496	13 Q918A2	Q918A2 gallus gall
16	101.5	5.1	588	3 Q9P982	Q9P982 lentinula e
17	101	5.1	503	13 Q9PUK7	Q9PUK7 anguilla an
18	101	5.1	1145	5 Q9T292	Q9T292 caenorhabdi
19	100.5	5.1	666	2 P96018	P96018 staphylococ

Query Match

100.0% Score 1979; DB 3; Length 599;

ALIGNMENTS

RESULT 1

ID Q10988 PRELIMINARY; PRT; 599 AA.
AC Q10988: P87339;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)
DE UV-ENONUCLEASE.
GN UVDE OR UVE1+.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SP972;
RX MEDLINE=96188860; PubMed=8614629;
RA Takao M., Yonemasa R., Yamamoto K., Yasui A.;
RT "Characterization of a UV endonuclease gene from the fission yeast
Schizosaccharomyces pombe and its bacterial homolog.";
RL Nucleic Acids Res. 24:1267-1271 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97175806; PubMed=9023111;
RA Davey S., Nass M.L., Ferrer J.V., Sidik K., Eisenberger A.,
RA Mitchell D.L., Freyer G.A.;
RT "The fission yeast UVDR DNA repair pathway is inducible.";
RL Nucleic Acids Res. 25:1002-1008 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Lyne M., Rajandream M.A., Barrell B.G., Lucas M., Gaillardin C.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; D78571; BAA11415.1;
DR EMBL; U78487; AAC49664.1;
DR EMBL; AL023859; CAA19577.1;
KW Endonuclease.
SQ SEQUENCE 599 AA; 68816 MW; C09ED7219D5895BE CRC64;

Q9sz77 arabidopsis
Q9yhd1 anguilla an
Q9lm14 arabidopsis
Q99ka3 mus musculu
Q52231 lactococcus
Q54680 lactococcus
Q9szn1 arabidopsis
Q9lqr5 arabidopsis
Q9spu5 zea mays (m
Q9ini9 arabidopsis
P94264 borrelia bu
Q99p78 mus musculu
Q66172 agrobacteri
Q9r488 agrobacteri
Q9esk9 mus musculu
Q54272 staphylococ
Q04650 arabidopsis
P73759 synechocyst
Q55455 synechocyst
Q99k16 mus musculu
Q99ne9 mus musculu
Q916w3 pseudomonas
Q9s6p3 chlamydia p
Q9z9g2 chlamydia p
Q81507 arabidopsis
Q9pmj1 campylobact

[illegible]

RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: TO ATPASE ALPHA/BETA CHAINS FAMILY.
 DR EMBL; AB016484; BAA36692.1; -;
 DR InterPro; IPR000194; ATPase_alpha_beta.
 DR InterPro; IPR003255; ATP_synt_B_C.
 DR Pfam; PF00006; ATP_synt_ab; 1.
 DR Pfam; PF01038; ATP_synt_B_C; 1.
 DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
 DR Hydrogen ion transport; Hydrolase.
 SQ SEQUENCE 509 AA; 56675 MW; 185BED824376451B CRC64;

Query Match 5.3%; Score 105; DB 5; Length 509;
 Best Local Similarity 17.8%; Pred. No. 6.7;
 Matches 79; Conservative 69; Mismatches 151; Indels 144; Gaps 18;
 QY 5 PREMFCDLKPWPGRGL-----GYACLTILSRMKRVF-----CSRT 43
 DB 132 PEDYLDIOGPIPKSRIYPEMIQTGISAITMNSIARGQKIPFSANGLPHNEIAAQI 191
 QY 44 CRITTIORGLSVKOLGTQNVLDLKLVEWNHNFHFMVRSVSDLPFPASHAKYGYTLE 103
 DB 192 CR-----QGGL---VKLPKQVMD-----SHEDNFAIV-----AAGVNM 225
 QY 104 ---FAQSHLEEVGKLANK-----YNHRUTMHPGOYTQIASPREVV----- 141
 DB 226 AARFFKSDFOEKGSDMNVCLFINLANDPTIERIITPRIALTAEFLAVQCEKHLVILTD 285
 QY 142 ---DSATRLD-AYHDEILSRMKLNEQLNKDAVLIHLGGTFEGKKEPFLDRFRKNYQRLS 196
 DB 286 MSSYAEALREVSAREEVPGRGPGCYMYTLATYERAGRVNGRNSITOI----- 337
 QY 197 DSVKARLVLENDVSWQDLPCLQELNIPVLVDWHHNIIVPGTLREGSLDMLPLIPTI 256
 DB 338 ---PILTPNDIDTHPIPLDTGYITEQIVIDRQLHNRQIYP-----PINVLPSSLRL 387
 QY 257 RETWTRKGIQKQH-----YSESADTAISGMRRAHSRDRVDFPPCDPTMDLMI--- 306
 DB 388 MKSATGEGTRKDHSDVSNQLYANTAIGKQVQAMKAVVGEAL-----TADMLYLE 439
 QY 307 -----EAKKEQAVF-----ELCRRYELQNPQCPLEIMGPEYDQTRGYPP 348
 DB 440 FLGKPEKTFIAQGPYENSRVSDSLNIGWELLRIF-----PKMKLRIPRNVINEYYP- 491
 QY 349 GAERKLTARKRSRKEEVEDEK 371
 DB 492 -----RKKVPSAQDKQEDOK 506

RESULT 12
 Q9GRIL PRELIMINARY; PRT; 881 AA.
 AC Q9GRIL;
 DT 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
 DE ND9 PROTEIN.
 GN ND9.
 OS Parametium tetraurelia.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
 OC Parametium.
 OX NCBI_TaxID=5888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Froissard M., Keller A.M., Cohen J.;
 RT "Nd9p, a novel protein with armadillo-like repeats involved in
 RT exocytosis: physiological studies using allelic mutants in
 RT Parametium."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ293945; CAC12829.1; -;
 SQ SEQUENCE 881 AA; 104104 MW; 8E27791A74CDD4E6 CRC64;

Query Match 5.3%; Score 104; DB 5; Length 881;
 Best Local Similarity 18.5%; Pred. No. 17;
 Matches 75; Conservative 61; Mismatches 126; Indels 144; Gaps 18;
 QY 6 REMFCDLKPWPGRGLGYACLTILSRMKRVFSCRTITTIORDG-----LESVKQL 60
 DB 44 KEGFDILE-----AKVSQCIDRLIVS-----SQDTYLFQIHFNKGRPKNVPLKNNI 90
 QY 61 GTQNVLDLKLVE--WNHNFGIHEMKV-----SSDL 89
 DB 91 HTLERDRLKFTQGWKTYMYKFLIEVRELPTYKGFERNYQVITMLKKPQVELMENQEM 150
 QY 90 FPFASHAKYGYTLEFAQSHLEEVGKLANKYHRLTMHPGOYT-QIASPREVVVDSAI--- 145
 DB 151 YKFNMHVNLGYNIFELNSY-----QNTKKGLYRGNQKSFQTLQVSETNELEVLEHMHNH 205
 QY 146 ROLAYHDEILSRMKLNEQLNKDAVLIHLGGTFEGKKEPFLDRFRKNYQRLSDSVKARLV 205
 DB 206 ODQIYTAEYVIVNALSEE--EKYWIHSAPIFK-----RKN-----L 240
 QY 206 ENDDVSWQDLPCLQELNIPVLVDWHHNIIVPGTLREGSLDMLPLIPTITRETWTRKGI 265
 DB 241 YNDLAEWK-----CWQTRAKVVM---KHODEP-----QOQTKKKR 272
 QY 266 TOKQHYSESADTAISGMRRAHSRDRVDFPPCDPTMDLMIKAEKEQAVFELCRYLEQ 325
 DB 273 TTKQATEME--YAVITMRKKYHPYLENF-----VDIVLTF-----LY 308
 QY 326 NPFCPLEIMGPEYDQTRGYYPGAEKRLTARKRSRKEEVEDEK 371
 DB 309 DPCKCKVEDNNEADQD-----QOELEEQLOESENEDGEEK 345
 RESULT 13
 Q9W6M4 PRELIMINARY; PRT; 502 AA.
 ID Q9W6M4;
 AC Q9W6M4;
 DT 01-NOV-1999 (TRENBLrel. 12, Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE V-TYPE ATPASE B SUBUNIT (EC 3.6.1.34).
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
 OX NCBI_TaxID=8022;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=GILL;
 RA Perry S.F., Beyers M.L., Johnson D.A.;
 RT "Cloning and molecular characterization of the trout (Oncorhynchus
 RT mykiss) vacuolar H⁺-ATPase B subunit";
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: TO ATPASE ALPHA/BETA CHAINS FAMILY.
 DR EMBL; AF140022; AAD33861.1; -;
 DR InterPro; IPR000194; ATPase_alpha_beta.
 DR InterPro; IPR003255; ATP_synt_B_C.
 DR Pfam; PF00006; ATP_synt_ab; 1.
 DR Pfam; PF01038; ATP_synt_B_C; 1.
 DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
 DR Hydrogen ion transport; Hydrolase.
 KW SEQUENCE 502 AA; 55738 MW; 0FC8CB6972D22BA0 CRC64;

Query Match 5.2%; Score 103; DB 13; Length 502;
 Best Local Similarity 19.4%; Pred. No. 9.3;
 Matches 54; Conservative 49; Mismatches 99; Indels 76; Gaps 11;
 QY 27 LNTILSRMKRVF-----CSRTCTITTIORGLSVKOLGTQNVLDLKLVEWNH 76
 DB 170 MNSIARGOKIPIFSAGLPHNEIAAQICR-----QAGL--VKK--SKDVMDYS-----DD 215
 QY 77 NFGIHEMRVSSDLPPFPASHAKYGYTLEFAQSHLEEVGKLANK-----YNH 121

```
Db 216 NFAIVFAAMGVNN-----ETARFKSDFEENGSMONVCLFLNLANDPITIERIITP 265
QY 122 RLTMHPGQVTOIASPREVV-----DSAIRDL-AYHDEILSRMKLNQKDAVLII 172
Db 266 RLALTSAYLAVQCEKHVILVILTDMSSYAEALREVSAREEVPGRGPGYMYTDLATIY 325
QY 173 HLGTFEGKKTLDPRKKNYQRLSDSVKARVLVLENDVSWSDVODLLPLCQELNIPVLVDW 232
Db 326 ERAGRVEGRNGSITQI-----PILTMPNDIDITHPIPDLTGYITEGOVYVDRQL 373
QY 233 HHNNIVPGTLREGSLDMLPIPTIRETWTKGITOQKH 270
Db 374 HNROIYP-----PINVLPSLSRLMKSAGEGMRKDH 405

RESULT 14
ID Q9ASE0 PRELIMINARY; PRT; 488 AA.
AC Q9ASE0;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DE PUTATIVE H+-TRANSPORTING ATP SYNTHASE.
GN P0456F08.19.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0456F08."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002901; BAB39419.1; -
SQ SEQUENCE 488 AA; 54279 MW; 99C6821732198EF1 CRC64;
```

```
Query Match 5.2%; Score 102.5; DB 10; Length 488;
Best Local Similarity 19.5%; Pred. No. 9.8;
Matches 73; Conservative 63; Mismatches 138; Indels 101; Gaps 16;

QY 27 LNTILSRMKERV-----CSRTCRTTTIQRDGLSVKQL-GTONVLDLKLVEWN 75
Db 154 MNSIARGOKIPLFSAAGLPHNEIAAQICR-----QAGL--VKRLEKSDNILESE---D 202
QY 76 HNFGIHMRVSSDLFPFASHAKYGYTLFAQSHLEEVG-----KLANK-----YN 120
Db 203 ENFAIVFAAMGVNN-----ETAOFKRDPEENGSMERVTFLNLANDPITIERIIT 252
QY 121 HRLTMHPGQVTOIASPREVV-----DSAIRDL-AYHDEILSRMKLNQKDAVLI 171
Db 253 PRIALTAYEYLAECGKHVILVILTDMSSYADALREVSAREEVPGRGPGYMYTDLATI 312
QY 172 IHLGTFEGKKTLDPRKKNYQRLSDSVKARVLVLENDVSWSDVODLLPLCQELNIPVLVD 231
Db 313 VERAGRIGRKGSIQI-----PILTMPNDIDITHPIPDLTGYITEGOVYVDRQ 360
QY 232 HHNNIVPGTLREGSLDMLPIPTIRETWTKGITOQKH-----YSESADPTAISGMK 284
Db 361 LHNROIYP-----PINVLPSLSRLMKSAGEGMRKDHSDVSNQLYANYAIGKDVQAMK 414
QY 285 RRAHSDRVDFPPCDPTMDLM-IEAKEKEQAVFELCRREYLN-----PPCPL 331
Db 415 AVVGEAL-----SSEDLLYLEFLDKFERKFTVQGAIDTRNIFOSLDLAWSLLRIFPR 467
QY 332 EIMGPEYDQTRDGY 346
Db 468 ELLHRIPAKTLDOY 482
```

```
RESULT 15
ID Q918A2 PRELIMINARY; PRT; 496 AA.
AC Q918A2;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE VACUOLAR H-ATPASE B SUBUNIT OSTEOCLAST ISOZYME (FRAGMENT).
GN VATB.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Andersen P., Lundberg L.G., Stone D.K.;
RT "Chicken vacuolar H-ATPase B subunit isozyme.";
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO ATPASE ALPHA/BETA CHAINS FAMILY.
DR EMBL; U61724; AAF73735.1; -
DR InterPro; IPR000194; ATPase_alpha_beta.
DR InterPro; IPR003255; ATP_synt_B_c.
DR Pfam; PF00006; ATP_synt_ab; 1.
DR Pfam; PF01038; ATP_synt_B_c; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
KW Hydrogen ion transport; Hydrolase.
FT NON_TER
SQ SEQUENCE 496 AA; 55049 MW; 1CC5D9FDA68BA7BA CRC64;
```

```
Query Match 5.1%; Score 101.5; DB 13; Length 496;
Best Local Similarity 19.2%; Pred. No. 12;
Matches 61; Conservative 55; Mismatches 111; Indels 91; Gaps 14;

QY 2 DHAP-----REMFDCLDKPIWGRGL-----GYAC---LNTILSRMKERV----- 39
Db 124 DRGPAVLAEDEFLDIMGQIPNPOCRIVPEPMIQTGISAIDGMNSIARGOKIPIFSAAGLPH 183
QY 40 ---CSRTCRTTTIQRDGLSVKQLGTONVLDLKLVEWNHNFIIHMRVSSDLFPFASHA 96
Db 184 NELAAQICR-----QAGL--VKK--SKVDYSE-----ENFAIVFAAMGVNN----- 222
QY 97 KYGYTLFAQSHLEEVGKLANK-----YNNHRLTMHPGQVTOIASPREVV 141
Db 223 ---ETARFKSDFEENGSMONVCLFLNLANDPITIERIITPRIALTAYEFLAYQCEKHVILV 279
QY 142 -----DSAIRDL-AYHDEILSRMKLNQKDAVLIHLGTFEGKKTLDPRKKNY 192
Db 280 ILTDMSSYAEALREVSAREEVPGRGPGYMYTDLATIVERAGRVEGRNGSITQI----- 335
QY 193 QRLSDSVKARVLVLENDVSWSDVODLLPLCQELNIPVLVDWHNNHNVPGTLREGSLDMLPL 252
Db 336 -----PILTMPNDIDITHPIPDLTGYITEGOVYVDRQLHNROIYP-----PINVLPS 381
QY 253 IPTIRETWTKGITOQKH 270
Db 382 LSRIMKSAIGEOMTRKDH 399
```

Search completed: January 15, 2002, 14:06:15
Job time: 954 sec

C-327, 384

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:50:09 ; Search time 120.01 Seconds

(without alignments)
357.991 Million cell updates/sec

Title: US-09-724-296-38

Perfect score: 3055

Sequence: 1 MCTTGLESLSLGRGAAPTV.....HKLNKYHDLWENHEKLSLS 580

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3055	100.0	580	AAW74473	Human MED1 endonuc
2	2958.5	96.8	585	AA144504	Human delta228-UV
3	2194.5	71.8	439	AA176548	Human ovarian tumo
4	1230	40.3	257	AAW88701	Secreted protein e
5	194	6.4	219	AAAG02051	Human secreted pro
6	148.5	4.9	326	AAAG16859	Arabidopsis thalia
7	148.5	4.9	327	AAAG16858	Arabidopsis thalia
8	148.5	4.9	352	AAAG16857	Arabidopsis thalia
9	145	4.7	1644	AAW25049	BRCA2 cancer suscep
10	145	4.7	3418	AAW19211	Human breast cance
11	145	4.7	3418	AAW23287	Human breast and o

12	145	4.7	3418	20	AAW04356	Human BRCA2 (Omi3)
13	145	4.7	3418	20	AAW04357	Human BRCA2 (Omi4)
14	145	4.7	3418	20	AAW04358	Human BRCA2 (Omi5)
15	145	4.7	3418	20	AAW04354	Human BRCA2 (Omi1)
16	145	4.7	3418	21	AAW77819	BRCA2 protein sequ
17	144	4.7	2329	18	AAW25038	Partial BRCA2 cance
18	143	4.7	3418	20	AAW04355	Human BRCA2 (Omi2)
19	141.5	4.6	782	11	AAW06991	Polypeptide antige
20	137	4.5	560	21	AAW19182	Lipid metabolism p
21	131.5	4.3	285	20	AAW14200	Mouse DNA demethyl
22	131.5	4.3	291	20	AAW14198	Human DNA demethyl
23	131	4.3	754	19	AAW33811	Tat stimulatory fa
24	130.5	4.3	411	20	AAW14197	Human DNA demethyl
25	130.5	4.3	411	22	AAW99915	Human protein sequ
26	130.5	4.3	411	22	AAW64314	Human protein H1.
27	130.5	4.3	411	22	AAW64844	Heart muscle cell
28	130	4.3	649	20	AAW90005	Expressed antigen
29	130	4.3	867	19	AAW98564	H. pylori GPO 175
30	130	4.3	1743	19	AAW98879	H. pylori GPO 175
31	127.5	4.2	281	20	AAW73829	Human prostate tum
32	127.5	4.2	281	20	AAW48439	Human prostate can
33	126	4.1	263	19	AAW4980	Human secreted pro
34	123	4.0	414	20	AAW14199	Mouse DNA demethyl
35	122.5	4.0	995	22	AAW06223	Peptide #4905 enco
36	122.5	4.0	1202	22	AAW17663	Peptide #4097 enco
37	121.5	4.0	1024	21	AAW42350	Arabidopsis thalia
38	121.5	4.0	1105	21	AAW42349	Arabidopsis thalia
39	121.5	4.0	1189	21	AAW42348	Arabidopsis thalia
40	120.5	3.9	2343	20	AAW80989	Canine factor VIII
41	120.5	3.9	2665	22	AAW14533	Peptide #967 enco
42	120.5	3.9	2665	22	AAW26950	Peptide #987 enco
43	120.5	3.9	2665	22	AAW02259	Peptide #941 enco
44	120.5	3.9	3266	21	AAW42491	Human ORF2255
45	119	3.9	646	21	AAW18188	Plasmodium falcipa

ALIGNMENTS

RESULT 1
 AAW74473
 ID AAW74473 standard; Protein: 580 AA.
 AC AAW74473;
 XX
 XX
 DT 19-MAY-1999 (first entry)
 XX
 XX
 DE Human MED1 endonuclease protein sequence.
 XX
 XX
 KW Endonuclease; MED1: human; methyl-CpG binding endonuclease-1;
 KW DNA fidelity; DNA manipulation; cancer; fragile X syndrome; therapy;
 KW myotonic dystrophy; Huntington's disease; spinocerebellar ataxia;
 KW Kennedy's disease; triplet repeat expansion disorder.
 XX
 XX
 OS Homo sapiens.
 XX
 XX
 PN WO9904626-A1.
 XX
 PD 04-FEB-1999.
 XX
 PF -28-JUL-1998; 98WO-US15828.
 XX
 PR 28-JUL-1997; 97US-0053936.
 XX
 PA (FOX-) FOX CHASE CANCER CENT.
 XX
 PI Bellacosa A;
 XX
 DR WPI; 1999-142462/12.
 DR N-PSDB; AAX22002*.
 XX
 PT New nucleic acid encoding human endonuclease MED1 involved in DNA mismatch repair - used for diagnosing susceptibility to cancer and

669 May 28 1999

PT fragile X syndrome, and therapeutically

PS Claim 8; Fig 3; 109pp; English.

XX This sequence is the human MED1 endonuclease of the invention.

CC MED1 (for methyl-CpG binding endonuclease-1) is used to screen for

CC specific modulators (potential therapeutic agents particularly mimetics

CC of MED1) and to study interactions involved in maintaining DNA fidelity,

CC for DNA manipulation and to raise antibodies. Susceptibility or

CC predisposition to cancer (particularly colorectal or endometrial,

CC especially hereditary non-polyposis colorectal cancer), or its prognosis,

CC where caused by alterations in the MED1-encoding gene, are identified by

CC sequence comparison, amplification, detecting altered polypeptide, and

CC restriction fragment mapping, hybridisation (particularly to probes

CC specific for a mutant allele). These same methods can also be used to

CC diagnose fragile X syndrome and other diseases (e.g. myotonic dystrophy,

CC Huntington's disease, spinocerebellar ataxia and Kennedy's disease)

CC associated with triplet repeat expansion. The DNA, or its fragments, are

CC used as probes and primers in the above diagnostic methods, also to

CC isolate homologous sequences, as sources of antisense sequences and for

CC gene transfer, particularly to restore drug sensitivity to drug-resistant

CC cancer cells.

XX Sequence 580 AA;

Query Match 100.0%; Score 3055; DB 20; Length 580;

Best Local Similarity 100.0%; Pred. No. 7 6e-284;

Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGTGLESLSGDRGAAPTTSERLVPDPNDLRKEDVAMELERVGEDEQMMIKRSSE 60

DB 1 mgttleslsigdrgaapttserslvpdpndlrkedvamelervgedeqmmikrsse 60

QY 61 CNPLLOEPITASQFAGTAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPOGLKFRS 120

DB 61 cnpllepiasagfagtagtecrksvpcgwerwvqrlfgktagrfdvfyispgglkfrs 120

QY 121 KSSLANYLHKNGETSLKPEDFTVLSKRGIKSRYKDCSMAALSHLQNSNNWNLR 180

DB 121 ksslanylhnkngetslkpedftvlskrgiksykdcsmalstshlqngsnwnlrrt 180

QY 181 RSKCKKDVMPSPSSSELSQESRGLSNFTSTHLLKKEDEGVDDVNFVRKPKGKVTILKG 240

DB 181 rskckkdvmppssselesqesrglsnftsthlkledgvddvnrkvrpkgkvtilk 240

QY 241 IPKTKTKGCRKSCSGFVQSDSKRESVCNKADAEPVAKQSOLDRTVCISDAGACGETL 300

DB 241 ipiktktkgcrkscsgfvqsdskresvcnkdasepvaqksqldrtvcisdagacgetl 300

QY 301 SVTSEENLVKKRSLSSGNGFCSEQKTSGLINKFCSAKDSEHNEYEDTFLESEIGT 360

DB 301 svtseenslvkkerslssgngfcseqktsglinkfcsakdsehneyedtfleseeigt 360

QY 361 KVEVVEREHLHTDILKKGSEMDNNSCPTKXDFTEKTFQEDTIPRQIERRKTSLVFSS 420

DB 361 kvevverehlhtdilkkgsemdnncsptrkdftektfqedtprqierrktslvfss 420

QY 421 KYNKEALSPPRKAPKFWTPRSPFNLVQETLFPHDPKWLTLATFLNFTSGKMAIPVLWK 480

DB 421 kynkealspprkapkfwtprrspfnlvqetlfphdpkwlltlatflnftsgkmaipvlwk 480

QY 481 FUEKYPSEAVARTADRWDSVSELLKPLGLYDLRAKTIKVSDEYLTWKWYPIELHGIGKY 540

DB 481 fuekypseavartadrdvswellkplglydlraktikvksdeyltwkwypielhigiky 540

QY 541 GNDSYRIFCVNEWKQVHPEDHLKNKYHDWLWENHEKISLS 580

DB 541 gndsyrifcvnewkqvhpedhlknkyhdwlwenhekisls 580

RESULT 2

AAY44504

ID AAY44504 standard; Protein; 565 AA.

XX AC AAY44504;

XX DT 27-MAR-2000 (first entry)

XX DE Human delta228-UV damage endonuclease.

XX KW delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;

XX KW glutathione-S-transferase signal peptide; uvcl+ gene product;

XX KW UV irradiation; DNA damage; UV radiation damage; photoproduct;

XX KW abasic site; platinum diaduct; mismatched nucleotide pairing;

XX KW nucleotide alkylation; skin cancer.

XX OS Homo sapiens.

XX PN WO9963828-A1.

XX PD 16-DEC-1999.

XX PF 08-JUN-1999; 99WO-US12910.

PR 08-JUN-1998; 98US-0088521.

PR 18-MAY-1999; 99US-0134752.

XX (UYEM-) UNIV EMORY.

XX PA Boetsch PW, Kaur B, Avery AM;

XX PI WPI; 2000-116417/10.

XX DR A new truncated ultraviolet damage endonuclease for treatment of skin

XX PT cancers -

XX PS Claim 16; Page 60; 133pp; English.

XX CC The present sequence is human delta228-UV damage endonuclease.

XX CC delta228-UVDE is a 288 amino acid deletion of the N-terminal of the

XX CC uvel+ gene product. This is expressed in frame with a GST leader sequence

XX CC to generate a fusion protein. This provides stable endonuclease fragments

XX CC for cleaving a double-stranded DNA molecule that has a distorted

XX CC structure resulting from UV radiation damage, a photoproduct, an abasic

XX CC site, mismatched nucleotide pairing, platinum diaduct, an intercalated

XX CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions

XX CC for internal or topical application and as a therapeutic agent for skin

XX CC cancers.

XX SX Sequence 565 AA;

Query Match 96.8%; Score 2958.5; DB 21; Length 565;

Best Local Similarity 97.4%; Pred. No. 1.3e-274;

Matches 565; Conservative 0; Mismatches 0; Indels 15; Gaps 1;

QY 1 MGTGLESLSGDRGAAPTTSERLVPDPNDLRKEDVAMELERVGEDEQMMIKRSSE 60

DB 1 mgttleslsigdrgaapttserslvpdpndlrkedvamelervgedeqmmikrsse 60

QY 61 CNPLLOEPITASQFAGTAGTECRKSVPCGWERVVKQRLFGKTAGRFDVYFISPOGLKFRS 120

DB 61 cnpllepiasagfagtagtecrksvpcgwerwvqrlfgktagrfdvfyispgglkfrs 120

QY 121 KSSLANYLHKNGETSLKPEDFTVLSKRGIKSRYKDCSMAALSHLQNSNNWNLR 180

DB 121 ksslanylhnkngetslkpedftvlskrgiksykdcsmalstshlqngsnwnlrrt 180

QY 181 RSKCKKDVMPSPSSSELSQESRGLSNFTSTHLLKKEDEGVDDVNFVRKPKGKVTILKG 240

DB 181 rskckkdvmppssselesqesrglsnftsthlkledgvddvnrkvrpkgkvtilk 240

QY 241 IPTKTKTKGCRKSCSGFVQSDSKRESVCNKADAEPVAKQSOLDRTVCISDAGACGETL 300

DB 241 ipiktktkgcrkscsgfvqsdskresvcnkdasepvaqksqldrtv----- 288

QY 301 SVTSEENSLVKKERSLSSGNSFCSEOKTSGIINKFCSAKDSHNEKYEIDTFLESEIGT 360
 Db 289 ---seenslvkkerslssgnsfscseqktsgiinkfcsakdsenhekyedtfleseeigt 345
 QY 361 KVEVVERKEHLHTDILKRGSEMNNCSPTKDKFTGKIFQEDTIPRTQIERRRKTSLYFSS 420
 Db 346 kvevverkehltidilkrsemnnncsptkdkftgkifqedtiprtqierrrktslyfss 405
 QY 421 KYNKEALSPRRKAFKWTTPRSPFNLVQETLFHDHPWKLLIATIFLNRSGKMAIPVLWK 480
 Db 406 kynkealspprrkafkwtpprspfnlvqetlfhdhpwklliatiflnrsgkmaipvlwk 465
 QY 481 FLEKPSAEVARTADRDVSELKPLGLYDLRAKTIVKFSDEYLTWKOWKYPTELHGIGKY 540
 Db 466 flekpsaeavartadrdvselkplgllydlraktivkfsdeyltkowkypielhgiky 525
 QY 541 GNDSYRIFCYNKQVHPDHKLKYNHDLWENHEKLSLS 580
 Db 526 gndsyrfcynkqvhpdklknynhdlwenheklsis (565)
 RESULT 3
 AAY76548 standard; Protein: 439 AA.
 XX AC AAY76548;
 XX DT 10-APR-2000 (first entry)
 XX DE Human ovarian tumor EST fragment encoded protein 44.
 XX KW Expressed sequence tag; EST; human; ovarian tumor; anticancer;
 XX KW gene therapy; treatment.
 XX OS Homo sapiens.
 XX PN DE19817557-A1.
 XX PD 21-OCT-1999.
 XX PF 09-APR-1998; 98DE-1017557.
 XX PR 09-APR-1998; 98DE-1017557.
 XX PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
 XX DR WPI; 1999-591920/51.
 XX DR N-PSDB; AAZ77469.
 XX PT New nucleic acid sequences expressed in ovarian, and some other, cancer
 XX PT tissues, and derived polypeptides, for treatment of ovarian cancer and
 XX PT identification of therapeutic agents .
 XX PS Claim 25; Page 261; 310pp; German.
 XX CC This invention describes novel nucleic acid (cDNA) sequences (A) which
 CC have anticancer activity and are highly expressed in ovarian tumor
 CC tissue (and some also in testis and breast cancer tissue). The products
 CC of the invention can be used for gene therapy. (A) are used (i) for
 CC recombinant expression of polypeptides (B) and (ii) to isolate complete
 CC genes. (B) are used (i) to identify agents suitable for treatment of
 CC ovarian cancer; (ii) directly for treating this form of cancer
 CC (including expression from gene therapy vectors) and (iii) for generation
 CC of specific antibodies. (A) are identified by assembling ESTs (expressed
 CC sequence tags) from a particular tissue type before comparison of
 CC expression patterns. This allows a significantly longer fragment of the
 CC gene to be revealed, so should reduce the number of failures associated
 CC with the fact that ESTs from different libraries may represent different
 CC parts of the same unknown gene, distorting the estimated frequency of
 CC occurrence in a particular tissue. AAY76505-y76638 represent protein

CC fragments encoded by the human ovarian tumor cDNA library derived EST
 CC fragments represented in AAZ77450-277572.
 XX Sequence 439 AA;

Query Match 71.8%; Score 2194.5; DB 20; Length 439;
 Best Local Similarity 96.0%; Pred. No. 1.6e-201;
 Matches 427; Conservative 3; Mismatches 6; Indels 9; Gaps 3;

QY 95 KQRLFGKTAGRED-VYFISPOGLKFRSKSSLANYLKNGETSLKPEDFDFTVLSKRGTKS 153
 Db 1 klllf--tsskfplisfsspgglkfrsksslanylkngetslkpedfdftvlskrgiks 58
 QY 154 RYKDCSMAALTSHLQNSNNNLRTSKCKKDVFMPPSSSELOESRGLSNFTSTHLL 213
 Db 59 rykdcsmaltshlqnqnnnnlrrtsckckdvfmppssseiqesrglsnftsthl 118
 QY 214 LKEDGCVDDVNFRRVKPKGVYTIKGIPIKTKKGRKSCSGFVQSDSKRESVCNKADA 273
 Db 119 lkedegvddvnnfrvkrpkgvtytikgipliktkkgrkscsgfvqsdskresvcnkada 178
 QY 274 ESEPVAKSQDLRTVCISDAGAGETLSVTSEENSLVKKERSLSSGNSFCSEQKTSGLI 333
 Db 179 esepvaqskqldrvtvcisdagagettlsvtseenslvkkerslssgnsfcseqktsgli 238
 QY 334 NKFCSAKDSHNEKYEIDTFLESEIGTKEVVERKEHLHTDILKRGSEMNNCSPTKDKF 393
 Db 239 nkfcsakdsenhekyedtfleseeigtkevverkehltidilkrsemnnncsptkdkf 298
 QY 394 TGEKTFQEDTIPRTQIERRRKTSLYFSSKYNKEALSPRRKAFKWTTPRSPFNLVQETLF 453
 Db 299 t-----edtiprtqierrrktslyfsskynkealspprrkafkwtpprspfnlvqetlf 352
 QY 454 HDPWKLLIATIFLNRSGKMAIPVLWKFLKYPSEAEVARTADRDVSELKPLGLYDLRA 513
 Db 353 hdpwklliatiflnrtsgkmaipvlwkflkypseavartadrdvselkplglydlra 412
 QY 514 KTVKFSDEYLTWKOWKYPTELHGIG 538
 Db 413 ktvksfdeyltkowkypielhgig 437

RESULT 4
 AAW88701 standard; Protein: 257 AA.
 XX AC AAW88701;
 XX DT 01-MAR-1999 (first entry)
 XX DE Secreted protein encoded by gene 168 clone HCFNFI1.
 XX KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX OS Homo sapiens.
 XX PN WO9854963-A2.
 XX PD 10-DEC-1998.
 XX PF 04-JUN-1998; 98WO-US11422.
 XX PR 18-DEC-1997; 97US-0070923.
 XX PR 06-JUN-1997; 97US-0048877.
 XX PR 06-JUN-1997; 97US-0048881.

PR 06-JUN-1997; 97US-0048884.
 PR 06-JUN-1997; 97US-0048893.
 PR 06-JUN-1997; 97US-0048896.
 PR 06-JUN-1997; 97US-0048899.
 PR 06-JUN-1997; 97US-0048915.
 PR 06-JUN-1997; 97US-0048949.
 PR 06-JUN-1997; 97US-0048964.
 PR 06-JUN-1997; 97US-0048972.
 PR 06-JUN-1997; 97US-0049020.
 PR 06-JUN-1997; 97US-0049375.
 PR 06-JUN-1997; 97US-0057628.
 PR 05-SEP-1997; 97US-0057635.
 PR 05-SEP-1997; 97US-0057644.
 PR 05-SEP-1997; 97US-0057647.
 PR 05-SEP-1997; 97US-0057650.
 PR 05-SEP-1997; 97US-0057661.
 PR 05-SEP-1997; 97US-0057667.
 PR 05-SEP-1997; 97US-0057761.
 PR 05-SEP-1997; 97US-0057764.
 PR 05-SEP-1997; 97US-0057770.
 PR 05-SEP-1997; 97US-0057775.
 PR 05-SEP-1997; 97US-0057778.
 PR 05-SEP-1997; 97US-0057784.
 PR 06-JUN-1997; 97US-0048875.
 PR 06-JUN-1997; 97US-0048878.
 PR 06-JUN-1997; 97US-0048882.
 PR 06-JUN-1997; 97US-0048885.
 PR 06-JUN-1997; 97US-0048894.
 PR 06-JUN-1997; 97US-0048897.
 PR 06-JUN-1997; 97US-0048900.
 PR 06-JUN-1997; 97US-0048916.
 PR 06-JUN-1997; 97US-0048962.
 PR 06-JUN-1997; 97US-0048970.
 PR 06-JUN-1997; 97US-0048974.
 PR 06-JUN-1997; 97US-0049373.
 PR 05-SEP-1997; 97US-0057629.
 PR 05-SEP-1997; 97US-0057634.
 PR 05-SEP-1997; 97US-0057642.
 PR 05-SEP-1997; 97US-0057645.
 PR 05-SEP-1997; 97US-0057648.
 PR 05-SEP-1997; 97US-0057651.
 PR 05-SEP-1997; 97US-0057662.
 PR 05-SEP-1997; 97US-0057668.
 PR 05-SEP-1997; 97US-0057762.
 PR 05-SEP-1997; 97US-0057765.
 PR 05-SEP-1997; 97US-0057771.
 PR 05-SEP-1997; 97US-0057776.
 PR 06-JUN-1997; 97US-0048876.
 PR 06-JUN-1997; 97US-0048880.
 PR 06-JUN-1997; 97US-0048883.
 PR 06-JUN-1997; 97US-0048892.
 PR 06-JUN-1997; 97US-0048895.
 PR 06-JUN-1997; 97US-0048898.
 PR 06-JUN-1997; 97US-0048901.
 PR 06-JUN-1997; 97US-0048917.
 PR 06-JUN-1997; 97US-0048963.
 PR 06-JUN-1997; 97US-0048971.
 PR 06-JUN-1997; 97US-0049019.
 PR 06-JUN-1997; 97US-0049374.
 PR 05-SEP-1997; 97US-0057627.
 PR 05-SEP-1997; 97US-0057634.
 PR 05-SEP-1997; 97US-0057643.
 PR 05-SEP-1997; 97US-0057646.
 PR 05-SEP-1997; 97US-0057649.
 PR 05-SEP-1997; 97US-0057654.
 PR 05-SEP-1997; 97US-0057656.
 PR 05-SEP-1997; 97US-0057760.
 PR 05-SEP-1997; 97US-0057763.
 PR 05-SEP-1997; 97US-0057769.
 PR 05-SEP-1997; 97US-0057774.
 PR 05-SEP-1997; 97US-0057777.
 PR 05-SEP-1997; 97US-0057784.

(HUMA-) HUMAN GENOME SCI INC.
 PA
 XX

PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA;
 PI Fan P, Feng P, Ferrite AM, Fischer CL, Florence C;
 PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW;
 PI Li Y, Moore PA, Ni J, Olsen JS, Rosen CA, Ruben SM;
 PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
 XX WPI: 1999-059865/05.
 DR N-PSDB; AAV84578.
 DR
 XX
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Claim 11; Page 564-565; 772pp; English.
 CC
 CC The invention relates to nucleic acid sequences (AAV84411 to AAV84633)
 CC encoding human secreted proteins (AAW8834 to AAW88756). The secreted
 CC protein gene sequences are deposited with the ATCC under deposit numbers
 CC ATCC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
 CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
 CC cells comprising recombinant vectors containing the nucleic acid
 CC sequences are used for the recombinant production of the secreted
 CC proteins. The polynucleotide and amino acid sequences are useful for are
 CC useful for preventing, treating or ameliorating medical conditions e.g.
 CC by protein or gene therapy. Pathological conditions can be also
 CC diagnosed by determining the amount of the new polypeptides in a sample
 CC or by determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the polynucleotides, based on
 CC which tissues they are most highly expressed in, and include developing
 CC products for the diagnosis or treatment of cancer, neurodegenerative
 CC disorders, developmental abnormalities and foetal deficiencies, blood
 CC disorders, tumours, leukemias, diseases of the immune system, autoimmune
 CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
 CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
 CC restenosis, prostate diseases, obesity, disorders involving osteoclasts
 CC such as osteoporosis, arthritis or malignancies, diseases of testes,
 CC lung or thymus, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.
 CC The present sequence represents human secreted protein (see descriptor
 CC line for gene number and clone identification).
 XX
 SQ Sequence 257 AA:

Query Match 40.3%; Score 1230; DB 20; Length 257;
 Best Local Similarity 93.1%; Pred. NO. 2.le-109;
 Matches 241; Conservative 3; Mismatches 9; Indels 6; Gaps 1;
 QY 160 MAALTSHLQNSNNWNLRTSKCKKDVFPSPSSSELQESRGLSNFTSTHLLKDEG 219
 DB 1 maaltshlqnsnnwnlrrtskckkdvfpmpssselqesrglsnftsthlkdeg 60
 QY 220 VDDVFRKVRKPKGVTKLGIPIKTKKGRKSCSGFVQSDSKRESVCNKADAESEPA 279
 DB 61 vddvfrkvrkpkgvtklkgipiktkkgrkscsgfvxdsdkresvcnkadaesepva 120
 QY 280 OKSOLDRTVCLSDAGACGTLSTVTEENSLVKKKERSLSSGNFCSEQKTSCLINKFCSA 339
 DB 121 qksqlartvcisdagacgtclstvtseenslvkkrerslssgnfcseqktsclinkfcsa 180
 QY 340 KDEHNEKYEDTFLESEETGKVEVVERKEHLHTDILKRGSEMDNNCSPTKRDKTGKIF 399
 DB 181 ksehnekyedtfleseeigtkvevverkehlhtdilkrgsemdnnccsptkrdktdft---- 235
 QY 400 QEDTTPRTQIERKTSLYP 418
 DB 236 -edtiprntdrkknkpvf 253

RESULT 5
 AAG02051
 ID AAG02051 standard; Protein: 219 AA.
 XX

RESULT 7
AAG16858
ID AAG16858 standard; Protein; 327 AA.
XX
AC AAG16858;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 17662.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-012180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139899.
PR 22-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

us-09-724-296-38.rag

Tue Jan-15 14:12:16 2002

PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139452.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 28-JUN-1999; 99US-0140991.
 PR 29-JUN-1999; 99US-0141287.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144684.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 4.9%; Score 148.5; DB 21; Length 352;
 Best Local Similarity 20.7%; Pred No. 1.9e-05;
 Matches 62; Conservative 47; Mismatches 102; Indels 89; Gaps 12;

XX Homo sapiens.
 OS WO9722689-A1.
 PN 26-JUN-1997.
 XX 17-DEC-1996; 96WO-US19598.
 XX 29-APR-1996; 96US-0639501.
 PR 18-DEC-1995; 95US-0573779.
 PR 20-DEC-1995; 95US-0575359.
 PR 21-DEC-1995; 95US-0576559.
 PR 11-JAN-1996; 96US-0585391.
 XX (HSCR-) HSC RES & DEV LP.
 PA (MYRI-) MYRIAD GENETICS I.
 PA (UYPE-) UNIV PENNSYLVANIA.
 PA (ENDO-) ENDO RECH INC.
 XX Couch F, Kamb A, Rommens JM, Simard J, Tavtigian SV;
 PI Weber BL;
 PI WPI; 1997-341680/31.
 DR N-PSDB; AAT69707.
 XX Human breast cancer susceptibility gene BRCA2 - useful for
 PT diagnosing breast cancer and screening for compounds to treat breast
 PT cancer
 XX Claim 1; Pages 90-106; 189pp; English.
 XX The present sequence is the human breast cancer susceptibility gene
 CC BRCA2 product, which can be used to diagnose breast cancer and
 CC screen for compounds to treat breast cancer. BRCA2 can also be used
 CC in gene therapy to restore wild type BRCA2 gene function to a cell,
 CC which has lost its or has altered (i.e. by virtue of a mutation in
 CC BRCA2) BRCA2 gene function.
 XX Sequence 3418 AA;
 SQ

Query Match 4.7%; Score 145; DB 18; Length 3418;
 Best Local Similarity 20.9%; Pred. No. 0.0017;
 Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;

Qy 36 KEDVAMELERVGEDEOMIKRSECNPLLOEPIASQAGTAGTCRKSVPCGWERVVK 95
 Db 972 ksdlsnidkipeknndymnkag-----llgplshsfsgsfirtsaskeikisehnikk 1026
 Qy 96 QRLFGKTAGRDYVIFISPOGLKFRSKSLANYLHKNGETSL-KPEDFDFTVLKRGIKSR 154
 Db 1027 skmffk-----dieeqyptsl---acveivntaldnqkklspqsin-tvsahlgssv 1077
 Qy 155 YKDCSMAALTSHL--NQSNNSNNLRTSKCKKDVMPSPSSSELSQESRGLSNFTSTHL 212
 Db 1078 vsdckshiltpqmlfskqdfnshnl-----tpsqkaeite-----lstill 1118
 Qy 213 LLKDEGVDDVNFRRKVRK-----PKGVVTILK----- 239
 Db 1119 -----eesgsqfefeqlkpsylqkstfevpenqmtlktseecdadlhvimnapsig 1174
 Qy 240 -----GIPKK-----TKGCRKSCSGFVOSDKRESVCNKADAESEPAOKSQL 284
 Db 1175 qvdskskfegteivkrfagllkndcnksasgyitde-----nevgrfgfysahgtkl 1227
 Qy 285 D-RVWCTSDACACCTGTSVTSEENSLVKKERSLSGNSFCSEQKTSIINKFCSAKDSE 343
 Db 1228 nvsteaigkavklfsdieniseets-----aevhplissskchd-----svvsmf---kien 1277
 Qy 344 HNEKYEDTFLSEEE-----IGTKVEYVERKEHLHTDILKRGSEMDNN---CSP 388
 Db 1278 hndk---tvseknkncqllilqnniemttgtfveei-----tenykrntenednkytaa 1327

Qy 389 TRK---DFTGEKIFQEDTIPRTOLERRKTSLYFSSKYN 423
 Db 1328 srnshnlfeqdskskndtv---clhkdetdlldtdqun 1363

RESULT 11
 AAW23287
 ID AAW23287 standard; Protein; 3418 AA.
 XX AC AAW23287;
 XX 21-FEB-1998 (first entry)
 XX Human breast and ovarian cancer susceptibility protein BRCA2.
 XX BRCA2; breast cancer; ovarian cancer; human;
 KW tumour suppressor protein; gene therapy; receptor.
 XX Homo sapiens.
 XX WO9730108-A1.
 XX 21-AUG-1997.
 XX 19-FEB-1997; 97WO-US03340.
 XX 20-FEB-1996; 96US-0603753.
 XX (UYVA-) UNIV VANDERBILT.
 XX (UNIW) UNIV WASHINGTON.
 XX Claire-King M, Holt JT, Jensen RA, Jetton TL, Page DL;
 PI Robinson-Benton CL, Szabo CI, Thompson ME;
 XX WPI; 1997-434733/40.
 DR N-PSDB; AAT84841.
 XX BRCA1 and BRCA2 tumour suppressor gene products - useful to inhibit
 PT breast and ovarian cancer cell growth and tumorigenesis, or treat
 PT gene linked hereditary or sporadic ovarian or breast cancer
 XX Claim 18; Page 72-88; 148pp; English.
 XX This protein comprises the human BRCA2 tumour suppressor protein
 CC that inhibits the growth of breast and ovarian cancer cells. BRCA2
 CC contains a granin, region, indicating that it is a secreted
 CC proteinin proteins. BRCA2 protein and DNA sequences encoding BRCA2
 CC (see AAT84841) can be used in the treatment of breast and ovarian
 CC cancers. Since the BRCA2 protein is secreted, it can be used to
 CC identify the BRCA2 receptor and hence to identify BRCA2 protein
 CC mimetics which act on the receptor for use in breast and ovarian
 CC cancer treatment. The BRCA1 gene (see AAT84840) and BRCA2 protein
 CC (see AAW23286) have also been characterised. Methods are claimed for
 CC the isolation of BRCA1 or BRCA2 receptors, and for treating or
 CC preventing breast and (gene-linked hereditary or sporadic) ovarian
 CC cancers using BRCA1 and BRCA2 proteins and genes.
 XX Sequence 3418 AA;
 SQ

Query Match 4.7%; Score 145; DB 18; Length 3418;
 Best Local Similarity 20.9%; Pred. No. 0.0017;
 Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;

Qy 36 KEDVAMELERVGEDEOMIKRSECNPLLOEPIASQAGTAGTCRKSVPCGWERVVK 95
 Db 972 ksdlsnidkipeknndymnkag-----llgplshsfsgsfirtsaskeikisehnikk 1026
 Qy 96 QRLFGKTAGRDYVIFISPOGLKFRSKSLANYLHKNGETSL-KPEDFDFTVLKRGIKSR 154
 Db 1027 skmffk-----dieeqyptsl---acveivntaldnqkklspqsin-tvsahlgssv 1077

PF 14-AUG-1998; 98WO-US16905.
 XX 22-MAY-1998; 98US-0084471.
 PR 15-AUG-1997; 97US-0055784.
 PR 07-NOV-1997; 97US-0064926.
 PR 12-NOV-1997; 97US-0065367.
 PR 01-MAY-1998; 98US-0071715.
 XX (ONCO-) ONCORMED INC.
 PA Eskandari T, Jackson GM, Murphy PD, Olson SJ, Park M;
 XX Rabin MB, Schryer B, White MB, Yoshikawa M;
 PI WPI: 1999-190163/16.
 DR N-PSDB; AAX30258.
 XX New coding sequence haplotypes of the human BRCA2 gene - used to
 PT develop products for determining susceptibility to, detection and
 PT treatment of breast or ovarian cancer
 XX Claim 29; Page 133-140; 226pp; English.
 PS The present invention describes genomic DNA which contains a BRCA2 gene
 CC where the first 12 nucleotides beginning exon 5 are 5'-TCCTGTTCTTCT-3',
 CC as in sequence (I) (see AAX30249), where nucleotides numbers
 CC 5782-5790 are GTTGTGTT as in sequence (IV) (see AAX30255), and where
 CC the last 20 nucleotides encoding exon 15 are 5'-CTGCGTGTCTCATAAACAG-3',
 CC as in sequence (II) (see AAX30251) and the first 20 nucleotides
 CC beginning exon 16 are 5'-CTGTATAGTATGCGTTC-3' as in sequence (III)
 CC (see AAX30253). Products and methods from the present invention can be
 CC used for identifying mutations in the BRCA2 gene leading to
 CC predisposition or higher susceptibility to breast or ovarian cancer. They
 CC can also be used for detection and gene therapy for breast and ovarian
 CC cancers. They can be used in methods for monitoring disease progression,
 CC for determining patients suited for gene and protein replacement
 CC progression, or for detecting the presence or quantifying the amount of a
 CC tumour growth inhibitor following such therapy. The BRCA2 protein,
 CC polypeptides, their functional equivalents, antibodies, and PNs may also
 CC be useful in the study of the characteristics of BRCA2 proteins, such as
 CC structure and function of BRCA2 in oncogenesis or subcellular
 CC localisation of BRCA2 protein in normal and cancerous cells. The present
 CC sequence represents human BRCA2 (oni4).
 XX Sequence 3418 AA;
 SQ

Query Match 4.7%; Score 145; DB 20; Length 3418;
 Best Local Similarity 20.9%; Pred. No. 0.0017;
 Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;
 QY 36 KEDVAMELERVGEDEQMIKRSSECNPLLOEPIASAOFCATAGTCRKSVPQWERYVK 95
 DB 972 ksdinidkipeknndymnkag-----llgpishstfggsfrtsanKeiklsehnikk 1026
 QY 96 QRLFGTAGTDFVYFISPOGLFKRSKSSLLANYLHKNGETSL-KPEDFDFTVLISKRGIKSR 154
 DB 1027 skmfk-----dieeqytsl---acveivtialdnqklskpsqin-tvsahlgssvv 1077
 QY 155 YKDCSMAALFTHL--NQSNNSNNLRTSKCKKVDVMPSSSELOEERGLSNFTSTHL 212
 DB 1078 vsdcknshilpqmlfskqdfnsnhl-----tpsqkaeite-----lslil 1118
 QY 213 LLKDEGVDDVNRKVRK-----PKGKVITLK----- 239
 DB 1119 ----eesgsgfctfrkpsylqkstfevpenqmtllktseecdadlhvimnapsig 1174
 QY 240 -----GIPIKK-----TKGCRKSGFGVQSDSKRESVCNKADAFSEPVAKSQL 284
 DB 1175 qvdskskfegteivkrkrfagllkndcnksasgyltde-----nevfgfgyahgtkl 1227
 QY 285 D-RTVCTSDACAGCETLUSVTSEENSLVKKERLSGGSNFCSEOKTSGIINKKCSAKDSE 343
 DB 1228 nvstealqkavklfsdieniseets---aevhplisisskchd-----svvsmf---kien 1277

QY 344 HNEKYEDTFLESEE-----ICTKVEVVERKEHLHTDILKRGSEMDNN--CSP 388
 DB 1278 hndk---tvseknnkcqllqnniemttgtfveei-----tenykrntenednkyytaa 1327
 QY 389 TRK----DFTGKIFQEDTIPRTQIERRTSLYFSSKYN 423
 DB 1328 srnshnleifdgsdsskndtv---cilkdetdltftdghn 1363
 RESULT 14
 AAY04358
 ID AAY04358 standard; Protein: 3418 AA.
 XX AAY04358;
 AC
 XX 21-JUN-1999 (first entry)
 DT Human BRCA2 (oni5) protein.
 DE
 XX Human; BRCA2; genetic testing; protein therapy; haplotype; detection;
 KW gene therapy; breast cancer; ovarian cancer.
 XX Homo sapiens.
 OS WO9909164-A1.
 PN 25-FEB-1999.
 PD 14-AUG-1998; 98WO-US16905.
 PP 22-MAY-1998; 98US-0084471.
 PR 15-AUG-1997; 97US-0055784.
 PR 07-NOV-1997; 97US-0064926.
 PR 12-NOV-1997; 97US-0065367.
 PR 01-MAY-1998; 98US-0071715.
 XX (ONCO-) ONCORMED INC.
 XX Eskandari T, Jackson GM, Murphy PD, Olson SJ, Park M;
 PI Rabin MB, Schryer B, White MB, Yoshikawa M;
 XX WPI: 1999-190163/16.
 DR N-PSDB; AAX30259.
 XX New coding sequence haplotypes of the human BRCA2 gene - used to
 PT develop products for determining susceptibility to, detection and
 PT treatment of breast or ovarian cancer
 XX Claim 20; Page 155-162; 226pp; English.
 PS The present invention describes genomic DNA which contains a BRCA2 gene
 CC where the first 12 nucleotides beginning exon 5 are 5'-TCCTGTTCTTCT-3',
 CC as in sequence (I) (see AAX30249), where nucleotides numbers
 CC 5782-5790 are GTTGTGTT as in sequence (IV) (see AAX30255), and where
 CC the last 20 nucleotides encoding exon 15 are 5'-CTGCGTGTCTCATAAACAG-3',
 CC as in sequence (II) (see AAX30251) and the first 20 nucleotides
 CC beginning exon 16 are 5'-CTGTATAGTATGCGTTC-3' as in sequence (III)
 CC (see AAX30253). Products and methods from the present invention can be
 CC used for identifying mutations in the BRCA2 gene leading to
 CC predisposition or higher susceptibility to breast or ovarian cancer. They
 CC can also be used for detection and gene therapy for breast and ovarian
 CC cancers. They can be used in methods for monitoring disease progression,
 CC for determining patients suited for gene and protein replacement
 CC progression, or for detecting the presence or quantifying the amount of a
 CC tumour growth inhibitor following such therapy. The BRCA2 protein,
 CC polypeptides, their functional equivalents, antibodies, and PNs may also
 CC be useful in the study of the characteristics of BRCA2 proteins, such as
 CC structure and function of BRCA2 in oncogenesis or subcellular
 CC localisation of BRCA2 protein in normal and cancerous cells. The present
 CC sequence represents human BRCA2 (oni5).
 XX Sequence 3418 AA;
 SQ

Search completed: January 15, 2002, 13:50:15
Job time: 304 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:45:11 ; Search time 120.01 Seconds
(without alignments)
511.063 Million cell updates/sec

Title: us-09-724-296-2

Perfect score: 4349

Sequence: 1 MTKLPILGKWKIKGLVQPTR.....KRLTARKRSRKEVEDEK 828

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4349	100.0	828	21 AAY44498	GST signal peptide
2	3092	71.1	600	21 AAY44500	GST signal peptide
3	1979	45.5	371	21 AAY44499	S. pombe delta228-
4	1259.5	29.0	564	16 AAR84565	Trypanosoma cruzi
5	1259.5	29.0	643	16 AAR84568	Trypanosoma cruzi
6	1237	28.4	229	21 AAY44501	Glutathione-S-tran
7	1221.5	28.1	472	16 AAR84569	Trypanosoma cruzi
8	1220	28.1	379	20 AAY31954	Glutathione-S-tran
9	1217.5	28.0	1140	16 AAR81731	GST-SEP fusion pro
10	1216.5	28.0	310	20 AAY31953	Glutathione-S-tran
11	1216.5	28.0	354	16 AAR84566	Trypanosoma cruzi

12	1212	27.9	824	21 AAY52288	Human HCP/GST fusi
13	1208.5	27.8	856	22 AAR70883	Expression vector
14	1207.5	27.8	362	16 AAR79948	G17 antigen. Chim
15	1206	27.7	426	16 AAR87091	Turkey prolactin/G
16	1205.5	27.7	579	21 AAY58676	GST-Cre recombinas
17	1204.5	27.7	324	22 AAB31613	Amino acid sequenc
18	1203	27.7	271	13 AAR21703	CAP-A-B fusion pro
19	1203	27.7	514	19 AAR47004	Glutathione-S-tran
20	1201.5	27.6	352	16 AAR79947	G15 antigen. Chim
21	1201	27.6	282	20 AAR74096	Glutathione-s-tran
22	1199.5	27.6	232	16 AAR72793	Amino acid sequenc
23	1199.5	27.6	245	22 AAB31612	GST/GI transport r
24	1199.5	27.6	247	20 AAW74114	Glutathione-s-tran
25	1199.5	27.6	247	20 AAW74120	GST/GI transport r
26	1199.5	27.6	247	20 AAW74103	GST/GI transport r
27	1199.5	27.6	248	20 AAW74112	GST/GI transport r
28	1199.5	27.6	248	20 AAW74113	GST/GI transport r
29	1199.5	27.6	248	20 AAW74121	GST/GI transport r
30	1199.5	27.6	248	20 AAW74122	GST/GI transport r
31	1199.5	27.6	248	20 AAW74100	GST/GI transport r
32	1199.5	27.6	248	20 AAW74101	GST/GI transport r
33	1199.5	27.6	248	20 AAW74102	GST/GI transport r
34	1199.5	27.6	249	20 AAW74115	GST/GI transport r
35	1199.5	27.6	257	20 AAW74111	GST/GI transport r
36	1199.5	27.6	257	20 AAW74119	GST/GI transport r
37	1199.5	27.6	257	20 AAW74124	GST/GI transport r
38	1199.5	27.6	257	20 AAW74126	GST/GI transport r
39	1199.5	27.6	257	20 AAW74105	GST/GI transport r
40	1199.5	27.6	257	20 AAW74109	GST/GI transport r
41	1199.5	27.6	258	20 AAW74117	GST/GI transport r
42	1199.5	27.6	258	20 AAW74104	GST/GI transport r
43	1199.5	27.6	259	20 AAW74130	GST/GI transport r
44	1199.5	27.6	259	20 AAW74118	GST/GI transport r
45	1199.5	27.6	259	20 AAW74125	GST/GI transport r

ALIGNMENTS

RESULT 1

AAV44498

ID AAY44498 standard; Protein; 828 AA.

XX AC AAY44498;

XX DT 27-MAR-2000 (first entry)

XX DE GST signal peptide and S. pombe UVDE fusion protein.

XX KW GST signal peptide; ultraviolet damage endonuclease; UVDE; Uvelp;

XX KW UV irradiation; DNA damage; UV radiation damage; fusion protein;

XX KW skin cancer; glutathione-S-transferase.

XX OS Schizosaccharomyces pombe.

XX OS Synthetic.

XX FH Key

XX FT Peptide

XX FT Protein

XX FT Misc-difference 11

XX FT Misc-difference 85

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

XX FT Misc-difference 147

SEQ ID NO: 2
IS AN OBLVIOUS CONST
RUCT

08-JUN-1999; 99WO-US12910.

XX 08-JUN-1998; 98US-0088521.
PR 18-MAY-1999; 99US-0134752.
XX (UDEM-) UNIV EMORY.
XX Doetsch PW, Kaur B, Avery AM;
PI WPI; 2000-116417/10.
DR N-PSDB; AA229858.
XX A new truncated ultraviolet damage endonuclease for treatment of skin
PT cancers -
XX Claim 7; Page 51; 133pp; English.

The present sequence is a fusion protein comprising the GST signal peptide and full length ultraviolet damage endonuclease (UVDE) from Schizosaccharomyces pombe. UVDE genes are isolated by introducing a foreign cDNA library into a repair-deficient E. coli strain and selecting for complemented cells by UV irradiation of the transformants. This provides stable endonuclease fragments for cleaving a double-stranded DNA molecule that has a distorted structure resulting from UV radiation damage, a photoproduct, an abasic site, mismatched nucleotide pairing, aplatium diaduct, an intercalated molecule or alkylation of a nucleotide. Uvelp can be used in compositions for internal or topical application and as a therapeutic agent for skin cancers.

XX Sequence 828 AA;
SQ

Query Match 100.0%; Score 4349; DB 21; Length 828;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 828; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKLPILGYKIKGLVQPRLLILEEYEEHLYERDEGDKWRNKKEFELGLEFPNLPYY 60
DB 1 mtklpilgykwikglvqprllileeyeehlyerdegdkwrnkkefelglefnpnlpyy 60
QY 61 IDGVKLTQSMATIRYADKHNHMLGCPKRAEISMLEGAVLDIRYGVSRAYSXDFTEL 120
DB 61 idgvkltsmaatiryadkhnmlgcpkeraeismllegavldirygvsriayskdfetl 120
QY 121 KVDFLSKLPKMFEDRLCHTKYLANGDVTHPDWFLXDALDVLVMDPMLCLDAPKLVLC 180
DB 121 kvdfslkplmkmfedrlchtkylngdvthpdlwflxdaldvlymdpmlcldapklvvc 180
QY 181 FKRIEAIPOIDKYLKSSKYIAWPLQGWQATEGGDHPKSDHLPVGRSMLRLKRNQI 240
DB 181 fkrleaiqidkylksskyiawplqgwqatfggdhpkshlvpgrsmrlrlkrnqi 240
QY 241 SKRIVFTILKQAKPGNHPCVPSCVITYSRFCHLPDTLKSLLPWSSKTTLSMLPQVNTG 300
DB 241 skrivftilkqakpgnhpcvpscvitysrfhclpdtlksllpwsstkttlsmlpqvntg 300
QY 301 ANSFSAETPVDLKKENETELANISGPHKSTSTSTRKARSSKKKATDSVSKIDESVAS 360
DB 301 ansfsaetpvdllkenetelansgphkstststrkarsskkkatdsvsdkidesvas 360
QY 361 YDSSTPLRRSRKPPVNNSSSESEEQISKATKKVQKEEEVEEVEDEKSLKNES 420
DB 361 ydsstplrrsrkppvnnssseseseqiskatkkvkqkeeeveeveedekslkness 420
QY 421 SDEPEPVVPEOLEPTISKRRSRSSAKNLEKSTMTNLDHAPREMFDCLDKPIPWGRIG 480
DB 421 sdepepvvpeoleptiskrrsrssaknlekestmtnldhapremfcdldkpiwgrig 480
QY 481 YACLNTILSRMKERVFCRTRITTIQDGLSVKQLGTQNVLDLKLKLVNHNHFGIHEM 540
DB 481 yacnltilsrmkervfcrtrittiqrldglsvkqlgtqnvldlklkvlvnhnfhgihem 540
QY 541 RYSSDLFPASHAKYGYTLFAQSHLEEVGKLANKYNHRLTMHPGOYTOITASPREVVVDS 600
DB 541 ryssdlfpashakgytlfashleevgklankynhrltmhpgoytoitasprrvvvds 600

Db 541 rvsdldfpashakgytlfashleevgklankynhrltmhpgoytiasprrvvvds 600
QY 601 AIROLAYHDEILSRMKLINEQNKDAVLIHLGOTFGCKKETLDRFRKNYQRLSDSVKARL 660
Db 601 airolayhdeilsrmlklineqnkdvlihlgtotfgckketldrfrknyqrlsdsvkarl 660
QY 661 VLENDVSWSVQDQLPLCQELNIPVLVDWHHNI VPTLREGSLDMLPIPTIRETWTRK 720
Db 661 vlendvswsvqddlplcqeinpvlvdwhhni vptlregslmliptiretwtrk 720
QY 721 GITOKQHYSESADPTAISGMKRAHSRDFPFCOPTMDLMEAEKEQAVFELCRRYE 780
Db 721 gitokqhyssadptaisgmkrhsrdfpfcopmtmdlmeaekqavfelfcrrye 780
QY 781 LONPPCPLEIMGPEYDQTRGCGYPPGAEKRLTARKRSRKEEVEDEK 828
Db 781 lonppcpleimgpeyqdrtdggyppgaeakrltarkrsrkeeeveedeek 828

RESULT 2
AA44500
ID AA44500 standard; Protein: 600 AA.
XX AA44500;
AC AA44500;
XX 27-MAR-2000 (first entry)
DT GST signal peptide and delta228 S. pombe UVDE fusion protein.
XX GST signal peptide; glutathione-S-transferase signal peptide;
KW ultraviolet damage endonuclease; UVDE; delta228-UVDE;
KW repair-deficient E. coli strain; UV irradiation; DNA damage;
KW UV radiation damage; photoproduct; abasic site; aplatium diaduct;
KW mismatched nucleotide pairing; nucleotide alkylation;
KW fusion protein; skin cancer.
XX Schizosaccharomyces pombe.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 11 /note= "Encoded by aaaa"
FT Misc-difference 85 /note= "Encoded by gtt"
FT Misc-difference 147 /note= "Encoded by gtt"
XX WO963828-A1.
XX
XX 16-DEC-1999; 99WO-US12910.
XX 08-JUN-1999; 99WO-US12910.
XX 08-JUN-1999; 98US-0088521.
PR 18-MAY-1999; 99US-0134752.
XX (UDEM-) UNIV EMORY.
XX Doetsch PW, Kaur B, Avery AM;
PI WPI; 2000-116417/10.
DR N-PSDB; AA229860.
XX A new truncated ultraviolet damage endonuclease for treatment of skin
PT cancers -
XX Disclosure; Page 56; 133pp; English.

The present sequence is a fusion protein comprising the GST signal peptide and delta228 truncated ultraviolet damage endonuclease (UVDE) from S. pombe. Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the S. pombe uvel gene product. This is expressed in frame with a GST leader sequence. Stable endonuclease fragments can be produced

for cleaving a double-stranded DNA molecule that has a distorted structure resulting from UV radiation damage, a photoproduct, an abasic site, mismatched nucleotide pairing, platinum diadduct, an intercalated molecule or alkylation of a nucleotide. Uvulp can be used in compositions for internal or topical application and as a therapeutic agent for skin cancers.

Query Match 71.1%; Score 3092; DB 21; Length 600;
Best Local Similarity 72.5%; Pred. NO. 5.6e-225;
Matches 600; Conservative 0; Mismatches 0; Indels 228; Gaps

QY	1	MTKLPIILGYWKIKGLVQPTRLLLILEYEEHLYERDEGDKWRNKKFELGLFPPMLPY	60
Db	1	mtklpilygwkwikglvqptrlllileeyeehlyerdegdkwrnkkfelglfppmlpy	60
QY	61	IDGVKLTQSMATIRIYIADKHNMGLGCPKRAEISMLEGAVIDIRYGVSRIVAKDFEFL	120
Db	61	idgvkltqsmaliyriadkhnmlgcpkeraeismlégavidiryvgsriayskafel	120
QY	121	KVDFLSKLPMLKMFEDRICHKTYTLNGDHVTHPDFMLYDALDVLVYMDPMCLDAFPKLYC	180
Db	121	kvdflsklpemlkmfedrlichktylngthvthpdfmlydaldvlymdpmcldafpklyc	180
QY	181	FKKRIEAIPOIDKYLKSSKYIAWPLQGWATCGGDDHPPKSOHLVPRGSMRLLLKKNIOI	240
Db	181	fkrieaipoidkylksskyiawplqgwatcggddhppksohlvprgsmrlllkknioi	229
QY	241	SKRIVFTILKQAKAFKNHPCVPSCVTITYSRFHCLDPTLKSLLPMSSKTTLSMLPQVNIG	300
Db	230	-----	229
QY	301	ANSFSAETPVDLKKENETELANTSGPHKKSTSTSTRKRARSSKKKATDVSODKIDESVAS	360
Db	230	-----	229
QY	361	YDSSTHLRSSRSRKPVNYSSESSESEQISKATKKVKQKEEEYVEVDEKSLKNSS	420
Db	230	-----	229
QY	421	SDSEFPVVPQLPTPSKRRRSRSSAKNLEKSTMNLDHAPREMPDCLDKPIPWGRGLG	480
Db	230	-----	252
QY	481	YACLNTILRSMKERVFCRSRCTRITTTQROGLSVKOLGTONVLDLKLVEWHNHNFIHPM	540
Db	253	yacintilsmkervfcsrctrittlqrdglesvkqlgtqnvldlklvewhnhfihi	312
QY	541	RVSSDLFPFASHAKYGYTLEFAQSHLEEVGKLANKYNHRLTWHPGGYTOIATSPREVVD	600
Db	313	rvssdlfpashakgytylefaqshleevgklankynhrltwhpggytqiasprevvds	372
QY	601	AIRDLAYHDEILSRMKLNQLNKDAVLIIHLGGTFEGKKETLDRFRKNYQORLSDSVKARL	660
Db	373	airdlayhdeilsrmlkneqlnkdvailhlgtfegkktldirfknqyrisdsvkarl	432
QY	661	VLENDVSVSVDLLPLCOELNPLVLDWHHNHNVPGTLREGSLDMLPIRTIETWTRK	720
Db	433	vlendvsvsvqdllplcqelnplvldwhhnlnvpgtlregslldmplirctretwtrk	492
QY	721	GITOKOHSADPTAISGMKRAHSDRVDFPPCPDPTMDLMEAKEBQAVFELCRRYE	780
Db	493	gitqkhsesadptaismkrrahsdrvdfppcpdptmdlmeakeqavfelcrrye	552
QY	781	LQNPCCPLEIMGPXYDOTRGVYPGCAEKLTARKRRSRKEEVEDEK	828
Db	553	lqnpccpleimgpyoqtrdgyppgaekrltarkrrsrkeevdeek	600
RESULT	3		
AAV	44499		

ID	AA144499	standard; Protein: 371 AA.
XX	AA144499	
XX	AA144499	
XX	AA144499	
DT	27-MAR-2000	(first entry)
XX	27-MAR-2000	(first entry)
DE	S. pombe	della228-uv damage endonuclease.
XX	S. pombe	della228-uv damage endonuclease.
KW	Delta228-UVDE;	ultraviolet damage endonuclease; GST signal peptide;
KW	glutathione-S-transferase	signal peptide; uvrl+ gene product;
KW	UV irradiation;	DNA damage; UV radiation damage; photoproduct;
KW	abasic site; platinum	diadduct; mismatched nucleotide pairing;
KW	nucleotide alkylation;	skin cancer.
XX	Schizosaccharomyces	pombe.
OS	Schizosaccharomyces	pombe.
XX	OS	Schizosaccharomyces pombe.
PN	W09963828-A1.	
XX	W09963828-A1.	
PD	16-DEC-1999.	
XX	16-DEC-1999.	
XX	08-JUN-1999;	99W0-US12910.
XX	08-JUN-1999;	98US-0088521.
PR	18-MAY-1999;	99US-0134752.
XX	18-MAY-1999;	99US-0134752.
XX	(UYEM-) UNIV EMORY.	
PA	(UYEM-) UNIV EMORY.	
XX	(UYEM-) UNIV EMORY.	
PI	Doetsch PW, Kaur B, Avery AM;	
XX	Doetsch PW, Kaur B, Avery AM;	
XX	WPI: 2000-116417/10.	
DR	WPI: 2000-116417/10.	
DR	N-PSDB: AAZ29859.	
XX	N-PSDB: AAZ29859.	
PT	A new truncated ultraviolet damage endonuclease for treatment of skin	
PT	cancers -	
XX	A new truncated ultraviolet damage endonuclease for treatment of skin	
PS	Claim 13; Page 53; 133pp; English.	
XX	Claim 13; Page 53; 133pp; English.	
CC	The present sequence is delta228-uv damage endonuclease. Delta228-UVDE is	
CC	a 288 amino acid deletion of the N-terminal of the S. pombe uvrl+ gene	
CC	product. This is expressed in frame with a GST leader sequence to	
CC	generate a fusion protein. This provides a stable endonuclease fragments	
CC	for cleaving a double-stranded DNA molecule that has a distorted	
CC	structure resulting from UV radiation damage, a photoproduct, an abasic	
CC	site, mismatched nucleotide pairing, platinum diadduct, an intercalated	
CC	molecule or alkylation of a nucleotide. Uvrlp can be used in compositions	
CC	for internal or topical application and as a therapeutic agent for skin	
CC	cancers.	
XX	cancers.	
SQ	Sequence 371 AA;	
Query Match 45.5%; Score 1979; DB 21; Length 371;		
Best Local Similarity 100.0%; Prod. No. 3e-141;		
Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	458	DDHAPRMFDCDLPIMWRGRIGYACLTNIIILSRMKERVFCSTCRITTTTORDGLESVKOL 517
Db	1	ddhaprmfcdcdlpimwrgrlgyacntililrsmkervfcsrctritttiqrdglesvkql 60
QY	518	GTONVLDLILKLVNHNHNGFIHMRVSSDLFPFASHAKYGYTLFEAQSHLEEVGKLANRYN 577
Db	61	gtqnvldliklvwnhngfihmrvssdlfpfashakgygtlfeaqshleevgklnkyn 120
QY	578	HLRLPMHFGQYTOIASPRVVDISAIRDLAYHDEILSRMKLNQLNKDAVLIHILGGTFEG 637
Db	121	hrlpmhfgqyqtqiasprvvvdaisairdlayhdeilsmklnqnlkdavliihggtfeg 180
QY	638	KKETLDRFRKNYQRLSDSVKARLVLENDVDSVSDLLPLCQELNIPLVLDWHHNHNVPG 697
Db	181	kketldrfrknyqrlsdsvkarlvlendvdsvsqdlplcqelnipvlvldwhhnivpg 240
QY	698	TLREGSLDMLPIPTIRETWTTRKGIOTKHYSADPTAISGMKRRASHDSRVDFPPCPDD 757

related nucleic acid and transformed cells

Disclosure; Page 40-41; 68pp; English.

AAR84565-R84569 are polypeptides of the TCR27 protein of T.cruzi. The proteins are all fusion products with glutathione-S-transferase (GST) and some contain a linker sequence. The TCR27 protein comprises a 95 amino acid (aa) N-terminal region; 69 repeats of a highly conserved 14 aa sequence, and a 68 aa C-terminal region. This sequence encodes the GST sequence, the Aq44 polypeptide contg. 16 of the 69 repeat units and also contains the amino and carboxy terminal peptides of TCR27. The TCR27 polypeptides of the invention are useful for the diagnosis of Chagas disease (American Trypanosomiasis), they are capable of detecting anti-T.cruzi antibodies; or for blood screening. The TCR27 protein has epitopes to which most T.cruzi infected individuals have antibodies. The TCR27 polypeptides will not react with serum from patients with leishmaniasis, schistosomiasis, or autoimmune disease and are hence less likely to cause false positives in diagnosis.

Sequence 643 AA;

Query Match 29.08; Score 1259.5; DB 16; Length 643;
Best Local Similarity 57.1%; Pred. No. 1.4e-86;
Matches 270; Conservative 55; Mismatches 103; Indels 45; Gaps 12;

Qy 5 PILGYWKIKGLVQPTRLLEYLEEKYEHLVERDEGDKWRNKKFELGLEPPLPYIDGD 64
Db 3 pilgywkikglvqptrllleyleekyeehlyerdegdkwrnkfflelgfelpnpyidgd 62
Qy 65 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRITAYSKDFETLKVDF 124
Db 63 vltqsmaliriyiadkhnmlggcpkeraeismlegavldirygvstriayskdfetlkvdf 122
Qy 125 LSKLPEMLKMPEDRLCHKTYLNGDHVTHPDMLYDALDVLYMDPMCLDAPFKLVCFKKR 184
Db 123 lsklpemlkmpedrclhcktylngdhvthpdmlydalvlymdpmlcldapfkvlvcfkr 182
Qy 185 IEALPQIDKYLKSSKYIAPLQGWQATFGGDHPKPSDHLVPRG--SMLRLKRNTOISK 242
Db 183 iealpoidkylksskylawplqgwqatfggdhpkpsd-lvprgspsqlqqaennitnsk 241
Qy 243 RIVFTILKQAFKGNHPCVSVCTIT-----YSRFHCLPD---TLKSLPLMSSKT 289
Db 242 k-entkirekvkkaekldalainratkleeernqaykaahkaeeekaktqrlittesen 300
Qy 290 -TLSMLPQVNI-----ANSFSAETPVDLKKENETELANISGPHKKSTSTSTKRRASSK 343
Db 301 inlkkrrpdavsnrdkkksetakt-d-evekqraaeakavetekqraaeat-kvaeaek 358
Qy 344 KKATDSV-----SDKIDSVASYDSSTHLRRSRKPKVNYNSSSESESESEEQISKATKKVK 399
Db 359 rkaaeakavetekqraaeatkvaeeakqkaaeakav-----etekqraaeatkva 410
Qy 400 QKEEEVEVEVDESKLNSSSEDFEPVPEQLETPISKRRRRSRSSAKNLEKE 452
Db 411 eekqraaeamkvaeeakqkaa-----eatkvaeeakqkaaeatkvaeeae 455

RESULT 6
AAY44501

ID AAY44501 standard; Protein; 229 AA.

AC AAY44501;

DT 27-MAR-2000 (first entry)

XX Glutathione-S-transferase signal peptide.

XX GST signal peptide; Glutathione-S-transferase signal peptide;
KW ultraviolet damage endonuclease; UVDE; delta228-UVDE;
KW UV irradiation; DNA damage; UV radiation damage; photoproduct;

abasic site; platinum diadduct; mismatched nucleotide pairing;
nucleotide alkylation; skin cancer.

Synthetic.

Key Location/Qualifiers
FT Misc-difference 11 /note= "Encoded by aaaa"
FT Misc-difference 85 /note= "Encoded by gtt"
FT Misc-difference 147 /note= "Encoded by gtt"
FT Misc-difference 147 /note= "Encoded by gtt"

W09963828-A1.

16-DEC-1999.

08-JUN-1999; 99WO-US12910.

08-JUN-1998; 98US-0088521.

18-MAY-1999; 99US-0134752.

(UYEM-) UNIV EMORY.

Doetsch PW, Kaur B, Avery AM;

WPL: 2000-116417/10.

N-PSDB; AA229861.

A new truncated ultraviolet damage endonuclease for treatment of skin cancers -

Claim 9; Page 58; 133pp; English.

The present sequence is GST signal peptide. Delta228, a truncated ultraviolet damage endonuclease from S. pombe uvel+ gene product is expressed in frame with the GST leader sequence to generate a fusion protein. This provides stable endonuclease fragments for cleaving a double-stranded DNA molecule that has a distorted structure resulting from UV radiation damage, a photoproduct, an abasic site, mismatched nucleotide pairing, platinum diadduct, an intercalated molecule or alkylation of a nucleotide. Uvelp can be used in compositions for internal or topical application and as a therapeutic agent for skin cancers.

Sequence 229 AA;

Query Match 28.4%; Score 1237; DB 21; Length 229;
Best Local Similarity 100.0%; Pred. No. 1.6e-85;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTKLPILGWYKIGLVQPTRLLEYLEEKYEHLVERDEGDKWRNKKFELGLEPPLPY 60
Db 1 mtklpilgywkikglvqptrllleyleekyeehlyerdegdkwrnkfflelgfelpnpy 60
Qy 61 IDGDVKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRITAYSKDFETL 120
Db 61 idgdvklqtsmaliriyiadkhnmlggcpkeraeismlegavldirygvstriayskdfetl 120
Qy 121 KVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDMLYDALDVLYMDPMCLDAPFKLV 180
Db 121 kvdfllsklpemlkmfedrclhcktylngdhvthpdmlydalvlymdpmlcldapfkvlvc 180
Qy 181 FKRRTEALPQIDKYLKSSKYIAPLQGWQATFGGDHPKPSDHLVPRGS 229
Db 181 fkkrrtealpoidkylksskylawplqgwqatfggdhpkpsdhlvprgs 229

RESULT 7

AAR84569

ID AAR84569 standard; Protein; 472 AA.

XX

AC AAR84569;
 XX 09-MAY-1996 (first entry)
 DT Trypanosoma cruzi TCR27 polypeptide, Ag8.
 XX
 DE TCR27; Chagas disease; repeat unit; diagnosis; blood screening;
 KW recombinant; fusion protein; glutathione-S-transferase.
 XX
 OS Trypanosoma cruzi.
 XX
 FH Location/Qualifiers
 FT Region
 FT /label= repeat_region
 FT /note= "15 of 69 repeat units of 14 amino acids"
 XX
 PN W09525797-A1.
 XX
 PD 28-SEP-1995.
 XX
 PF 20-MAR-1995; 95WO-0503191.
 XX
 PR 24-MAR-1994; 94US-0216894.
 XX
 PA (KIRC/) KIRCHHOFF L V.
 PA (OTSU/) OTSU K.
 XX
 PI Kirchhoff LV, Otsu K;
 XX
 DR WPI: 1995-344618/44.
 DR N-PSDB; AAT05333.
 XX
 XX New polypeptide(s) from the TCR27 protein of Trypanosoma cruzi - as
 PT immunoassay reagent for specific diagnosis of Chagas disease, also
 PT related nucleic acid and transformed cells
 XX
 PS Disclosure; Page 44-45; 68pp; English.
 XX
 CC AAR84565-R84569 are polypeptides of the TCR27 protein of T. cruzi
 CC The proteins are all fusion products with glutathione-S-transferase
 CC (GST) and some contain a linker sequence. The TCR27 protein comprises
 CC a 95 amino acid (aa) N-terminal region; 69 repeats of a highly
 CC conserved 14 aa sequence and a 68 aa C-terminal region. This sequence
 CC encodes the GST sequence, the Ag8 polypeptide contg. 15 of the 69
 CC repeat units and also contains a linker sequence. The TCR27
 CC polypeptides of the invention are useful for the diagnosis of Chagas
 CC disease (American trypanosomiasis), they are capable of detecting
 CC anti-T. cruzi antibodies; or for blood screening. The TCR27 protein
 CC has epitopes to which most T. cruzi infected individuals have
 CC antibodies. The TCR27 polypeptides will not react with serum from
 CC patients with leishmaniasis, schistosomiasis, or autoimmune disease
 CC and are hence less likely to cause false positives in diagnosis.
 XX
 SQ Sequence 472 AA;

Query Match 28.1%; Score 1221.5; DB 16; Length 472;
 Best Local Similarity 56.7%; Pred. No. 6.6e-84;
 Matches 255; Conservative 54; Mismatches 110; Indels 31; Gaps 9;

QY 5 PILGWYKIGLVQPTLLLEYLEEKYEHLHYERDEGDKWRNKKFELGLFPPNLPYIDGD 64
 DB 3 PILGWYKIGLVQPTLLLEYLEEKYEHLHYERDEGDKWRNKKFELGLFPPNLPYIDGD 62
 QY 65 VKLTQSMATIRYTDKHNMLGGCPKRAEISMLEGAVLDIRYGVSRVAYSKDFETLKVD 124
 DB 63 VKLTQSMATIRYTDKHNMLGGCPKRAEISMLEGAVLDIRYGVSRVAYSKDFETLKVD 122
 QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHTVDFDMLYDALDVLVYMDPNCMLDAFPLKVCFFKR 184
 DB 123 LSKLPEMLKMFEDRLCHKTYLNGDHTVDFDMLYDALDVLVYMDPNCMLDAFPLKVCFFKR 182
 QY 185 IEATPQIDKYLKSSKYTAWPLOGMQATFGGDDHPKSDHLVPRGSMRLRLKRNLIQISKRI 244

DB 183 IEATPQIDKYLKSSKYTAWPLOGMQATFGGDDHPKSDHLVPRGSMRLRLKRNLIQISKRI 237
 QY 245 VFTILKQKAFKGNHPCVPSVCTITYSRFHCLEDTLKSLLPMSKTTLSMLPQVNTIGANSF 304
 DB 238 a-eaekqkaaeat-----kvaeeakqraaeatk--vaeekqkaaeatkvaeeakqr 286
 QY 305 SAETP--VDLKKENETELANISGPHKKSSTSTRKARSKKKATDSVSDKIDESVASVD 362
 DB 287 aaeatkvaeeakqkaaeatkvagdekqkaaeat-kvaeeakqkaaeat--kvae--aekq 341
 QY 363 SSTHLRRSSRSKPKPVNYNSSESESEEQISKATKKYKQKEEYVEEYDEKSLKNESSD 422
 DB 342 kaeeatkvaeeakqkaaeatkvaeeakqkaaeatkvaeeakqkaaeatkvaeeakqkaa- 400
 QY 423 EFEPVVPQLETPISKRRRSRSKAKNLEKE 452
 DB 401 -----eatkvaeeakqkaaeatkvaee 423

RESULT 8
 AAY31954
 ID AAY31954 standard; Protein: 379 AA.
 XX
 AC AAY31954;
 XX
 DT 21-DEC-1999 (first entry)
 XX
 DE Glutathione-S-transferase:Weel(11-152) fusion protein.
 XX
 KW GST:Weel(11-152); glutathione-S-transferase; GST: Weel; human;
 KW mitosis; cell cycle checkpoint; cyclin-dependent kinase; Cds1.
 XX
 OS Chimeric - Schistosoma mansoni.
 OS Chimeric - Homo sapiens.
 XX
 PN W09949061-A1.
 XX
 PD 30-SEP-1999.
 XX
 PF 26-MAR-1999; 99WO-US06704.
 XX
 PR 27-MAR-1998; 98US-0079752.
 XX
 PA (SCRI) SCRIPPS RES INST.
 XX
 PI Russell P, Boddy MN, Funari B;
 XX
 DR WPI; 1999-591103/50.
 DR N-PSDB; AAZ20067.
 XX
 PT Truncated fusion proteins and related polynucleotides useful for
 PT screening for compounds which modulate weel function
 XX
 PS Claim 6; Page 40-42; 46pp; English.
 XX
 CC The present sequence represents a fusion protein, termed
 CC GST:Weel(11-152), comprising glutathione-S-transferase and a
 CC truncated weel protein composed of amino acid residues 11-152 of
 CC native weel. The fusion protein was expressed in Escherichia
 CC coli cells that had been transformed with a pCEX vector into which
 CC a PCR fragment encoding amino acids 11-152 of weel had been cloned.
 CC Weel participates in the regulation of cell cycle checkpoints during
 CC mitosis by catalysing phosphorylation of the cyclin-dependent
 CC kinase that initiates mitosis, Cdc2. Weel is itself phosphorylated
 CC by Cdc1. The fusion protein is useful in assays for evaluating weel
 CC function and for screening for modulators of weel and Cds1. It can
 CC be used in combination with other components of the cell cycle,
 CC particularly Cds1, Chk1, Cdc2, Cdc25 and other checkpoint proteins,
 CC in methods to identify and characterise regulatory pathways and to
 CC identify compounds which modulate components in the pathways,
 CC particularly Cds1.

SQ Sequence 379 AA;

Query Match 28.1%; Score 1220; DB 20; Length 379;
 Best Local Similarity 60.3%; Pred. No. 6.2e-84;
 Matches 252; Conservative 32; Mismatches 50; Indels 84; Gaps 8;

QY 5 PILGYWKIKGLVQPTLLILEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
 DB 3 pilgywkikglvqptllileeyeehlyerdegdkwrnkkfelglefnpnlpyyidgd 62

QY 65 VKLTQSMALIIRYIADKHNLGCGPKRAEISMLEGAVLDIRYGVSRAYSXKDFETLKVDF 124
 DB 63 vltqsmaliiryiadkhnlgcgpkeraeismlegavldirygvsrayskdfetlkvdf 122

QY 125 LSKLPEMLKMFEDRLCHTKYLNQDHTVHPDMLYDALDVLVYMDPMLCLDAPPKLVCFKKR 184
 DB 123 lsklpemlkmfedrclchktylnqdhvthpdmlydaldvlymdpmlcldafpklvcfkkr 182

QY 185 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDHPPKSDHLVPRGSM----- 230
 DB 183 ieapqidkylksskyiawplqgwqatfggdhppksd-lvprgspgisgggghmrsv 241

QY 231 -LRLKRNQIKSRIVFTILKQKAFKGNHPCVSVCTITYSRFHCLPDTLKSLPLMSSTK 289
 DB 242 glrrsqrsmlnra---cll-----aptpsslydannst 273

QY 290 TSLMPLQVNICANSFSAETPYDLKKENETELANISGPHKKSTS----- 332
 DB 274 -----sstseqp-----ntsfslfgrkgttspsfsfshaaphlpsps 314

QY 333 -TSTRKARSKKATKATSVSK-----IDESVASYDSSTHLRNRSSRKKPVNYSSESE 386
 DB 315 fthsqpqiqaqvprpslfdprnlvrsrsrldgspslpsvaqvanpihtapspsd 372

RESULT 9
 AAR81731 ID AAR81731 standard; Protein; 1140 AA.
 AC AAR81731;
 XX
 DT 29-MAY-1996 (first entry)
 XX
 DE GST-SEP fusion protein fragment.
 XX
 KW Glutathione-S-transferase; GST; FK506 binding protein; FKBP; PCR;
 KW rapamycin; FKBP-rapamycin binding protein; Molt4 cell; amplify;
 KW fusion protein; GST-FKBP12; immunomodulatory agent; primer;
 KW antitumour agent; detection; antisense DNA; immune system.
 XX
 OS Homo sapiens.
 XX
 OS EP676471-A2.
 PN
 PD 11-OCT-1995.
 XX
 PF 07-MAR-1995; 95EP-0301475.
 XX
 PR 13-FEB-1995; 95US-0384524.
 PR 08-MAR-1994; 94US-0207975.
 PR 26-SEP-1994; 94US-0312023.
 XX
 PA (AMHP) AMERICAN HOME PROD CORP.
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 PI Caggiano TJ, Chen Y, Failli AA, Molnar-Kimber KL;
 PI Nakanishi K;
 XX
 DR WPI; 1995-346091/45.
 DR N-PSDB; AAT00771.
 XX
 PT New effector proteins of rapamycin - which bind to a

PT glutathione-S-transferase-FK506 binding protein-rapamycin complex
 XX
 PS Example 2; Page 29-35; 44pp; English.
 CC This sequence represents a fragment of a fusion protein between
 CC glutathione S transferase (GST) and sirolimus effector protein (SEP)
 CC gene. It represents the fragment beginning at the linker sequence
 CC between the GST and SEP45 fragment. The cDNA encoding this sequence
 CC was isolated from Molt 4 human T-cell leukaemia cells (ATCC CRL 1582).
 CC The sep gene was amplified in five fragments using the primers given in
 CC AAT00756-69. The amplified SEP gene was inserted into a vector already
 CC containing the GST gene and the fusion protein expressed. The fusion
 CC protein was used in the isolation of a protein of mammalian origin
 CC which binds a glutathione-S-transferase (GST)-FK506 binding protein
 CC (FKBP)-rapamycin complex. The FKBP-rapamycin binding protein and
 CC corresponding DNA was isolated from Molt4 cells using a complex of the
 CC fusion protein GST-FKBP12 and rapamycin. The isolated proteins have
 CC molecular weights of 125, 148, 208 and 210 kD. They can be used for
 CC identifying an immunomodulatory, or an antitumour agent. They can also
 CC be used in the detection of rapamycin, rapamycin analogues or
 CC metabolites when complexed with FKBP. Antisense DNA can be used to
 CC modulate the immune system of a mammal.
 XX
 SQ Sequence 1140 AA;

Query Match 28.0%; Score 1217.5; DB 16; Length 1140;
 Best Local Similarity 36.2%; Pred. No. 4.8e-83;
 Matches 344; Conservative 103; Mismatches 261; Indels 241; Gaps 34;

QY 5 PILGYWKIKGLVQPTLLILEEYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
 DB 3 pilgywkikglvqptllileeyeehlyerdegdkwrnkkfelglefnpnlpyyidgd 62

QY 65 VKLTQSMALIIRYIADKHNLGCGPKRAEISMLEGAVLDIRYGVSRAYSXKDFETLKVDF 124
 DB 63 vltqsmaliiryiadkhnlgcgpkeraeismlegavldirygvsrayskdfetlkvdf 122

QY 125 LSKLPEMLKMFEDRLCHTKYLNQDHTVHPDMLYDALDVLVYMDPMLCLDAPPKLVCFKKR 184
 DB 123 lsklpemlkmfedrclchktylnqdhvthpdmlydaldvlymdpmlcldafpklvcfkkr 182

QY 185 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDHPPKSDHLVPRGSMRLKLNKIQISKRI 244
 DB 183 ieapqidkylksskyiawplqgwqatfggdhppksd-lvprg-----spdisggg 234

QY 245 VFTILKQK-AFKGNHPCVPSVCTITYSRFHCLPDTLKSLPLMSSTTL-----SMLPQVN 298
 DB 235 gggildsdmsfkyaslcgks-----grlalehktlrvlllgvdpsrqldhplptvhpqvt 288

QY 299 IG--ANSPSAETPVD-----LKKENETELANISGPHK----- 328
 DB 289 yaymkumwksarkidafqhmghfvqtmgqgaahatdqghkqlhklmarcfklkgew 348

QY 329 ----KSTSTSTRKRARSKKATDSVSKIDES-----VASVDSSTHLRNRSSR--- 372
 DB 349 qinlqqineestipkvlygysaateh-----drswykwahawamnfavllhkhqndprd 403

QY 373 SKKPVNYSSESESESEQISKATKKVKQKEEYVEEVDKSLKNESSDEFEPPVPEQL 432
 DB 404 ekkkrlrhasga-----nitnattaattaattastegsnseaeastensp-tpspl 456

QY 433 ETPISK-----RRSRSSAKNLE---KSTMLND-DHAP----- 462
 DB 457 qkkvtedlsktilmytvpavqgffrsislsrgunlqdtlrvltlwdyghwpdvneale 516

QY 463 -----REMFCDLDPKIPWRGRLGYACLTNLTLSMKRVCSCRTCC---RIIT 505
 DB 517 gvkaiqidtwlqvipqliaridprlvgrlihqldtdigrvhpqaliplvtvasksttt 576

QY 506 IORDGLSVKQLGTQNVLDLKLVEVNNHNFGLHFWRSVSDLPFPFAS--HAKYGYTLEPAQ 563
 DB 577 arhnaankl-----lkmcelhsntlvqgammvseelirvailwh-----emhw 619

	QY	564	SHLBEVKGLANKYNHRLTWHHPGQYTQIASPREVVVDSAIRDL-----AYHDEILSRMK	610
	DB	620	egleearsl--yfgernv--kgmf-evlieplhammergpgqtiketstfnqaygrdimeage	674
	QY	617	LNEQLNKDAVLIITHLGSTFGKKEITLDRFKNKYORLSDSVKARLVLENDVDWSVSODLLP	676
	DB	675	wcrkymk-----sgnvkdltqawdlvyhvfriskqlpqtlstlelqvys-----pkll	722
	QY	677	LCOELNIPVLVDWHHNHVPGTL--REGSLDMLPIPTIRETWTKIGTOKKHYSE----	730
	DB	723	mcrdlel-----avpgtydpnqpilriqiapslv-----itskqrprkltim	766
	QY	731	-----SADPTAIS---GMKRRA-----HSD	747
	DB	767	gsnghgefvlkghedlrqdervmqlglvnltilandptsrknlsiqryaviplstnsq	826
	QY	748	RVFDFPCDPMTDLMIETAKEGEAVFELCRRYEQLNPCCPLEIMPEYD	796
	DB	827	ligwvhpcdtlhaldyrekkllniehrmlt-----mapdyd	867
		RESULT	10	
		AAAY31953		
		ID	AAAY31953 standard; Protein: 310 AA.	
		XX	AC AC	
		AAAY31953;		
		21-DEC-1999	(first entry)	
		XX	Glutathione-S-transferase:Weel(11-72) fusion protein.	
		DE		
		XX	GST:Weel(11-72); glutathione-S-transferase; GST; Weel; human;	
		KW	mitosis; cell cycle checkpoint; cyclin-dependent kinase; Cdsl.	
		KW		
		OS	Chimeric - Schistosoma mansoni.	
		OS	Chimeric - Homo sapiens.	
		PN	WO9949061-A1.	
		PD	30-SEP-1999.	
		PF	26-MAR-1999; 99WO-US06704.	
		PR	27-MAR-1998; 98US-0079752.	
		PA	(SCRI) SCRIPPS RES INST.	
		XX		
		PI	Russell P, Boddy MN, Funari B;	
		DR	WPI; 1999-591103/50.	
		DR	N-PSDB; AAZ20066.	
		XX	Truncated fusion proteins and related polynucleotides useful for	
		PT	screening for compounds which modulate Weel function	
		XX		
		PS	Claim 5; Page 37-38; 46pp: English.	
		XX		
		CC	The present sequence represents a fusion protein, termed	
		CC	GST:Weel(11-72), comprising glutathione-S-transferase and a	
		CC	truncated Weel protein composed of amino acid residues 11-72 of	
		CC	native weel. The fusion protein was expressed in Escherichia	
		CC	coli cells that had been transformed with a pGEX vector into which	
		CC	a PCR fragment encoding amino acids 11-72 of Weel had been cloned.	
		CC	Weel participates in the regulation of cell cycle checkpoints during	
		CC	mitosis by catalysing phosphorylation of the cyclin-dependent	
		CC	kinase that initiates mitosis, Cdc2. Weel is itself phosphorylated	
		CC	by Cdc1. The fusion protein is useful in assays for evaluating Weel	
		CC	function and for screening for modulators of Weel and Cds1. It can	
		CC	be used in combination with other components of the cell cycle,	
		CC	particularly Cds1, Chkl, Cdc2, Cdc25 and other checkpoint proteins,	
		CC	in methods to identify and characterise regulatory pathways and to	
		CC	identify compounds which modulate components in the pathways,	

```

CC      particularly CcdL.
XX
SQ      sequence      310 AA:

Query Match      28.0%; Score 1216.5; DB 20; Length 310;
Best Local Similarity 69.7%; Pred. No. 8.6e-84;
Matches 244; Conservative 18; Mismatches 25; Indels 63; Gaps

QY      5 PILGYWKIKGLVOPTRLLEYLLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db      3 pilgywkikglvqptrllileyleekyeelhyerdegdkwrnkkfelglefnpnyidgd 62

QY      65 VKLTQSMAIIRYIADKHNMGLGGCPKERATISMLEGAVLDIRYGVSRIAYSKDFETLKVD 124
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db      63 vltqsmairiyiadkhnmglgpceraeismlegavldirygvsrriayskdfetlkvd 122

QY      125 LSKLPEMLKMFEDRLCHKIYTLNGDHYTHDFPMLYDALDVVLVMDPMLCLDAFPKLVCFKKR 184
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db      123 lsklpemlkmfedrlchkkytlngdhythpdmlydalgvilymdpmcldafpklvcfkkr 182

QY      185 IEAIPQIDKYLKSSKIYAWPLOGWQATFGGDHPKPSDHLVPRGSM----- 230
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db      183 ieaipqidkylksskiyawlpgwqatfggdhpkpsd-lvrgspgsgggggghmrtsy 241

QY      231 -LRLKRNQISIRIVFTILKQAFKGNHPVPSVCTIYSRPHCLPDPLTKLLPMSSKT 289
      ||| ||| :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db      242 qlrksrsmalnra--tll-----appcpsslydannst 273

QY      290 TSLMLQPVNIGANSFSAETPVDLKKENETELANISGPHKKSTSTSRKRA 339
      ||| ||| :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      ||| ||| :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db      274 -----sstsgkp-----ntsfllfqprkqttspsfsha 304

```

AA AAR84565-R84569 are polypeptides of the TCR27 protein of T.cruzi
CC The proteins are all fusion products with glutathione-S-transferase
CC

RESULT 13
 AAB70883
 ID AAB70883 standard; Protein; 856 AA.
 AC AAB70883;
 XX
 DT 12-JUL-2001 (first entry)
 XX
 DE Expression vector pGEX-4T-1-Amsu5 containing amylosucrase.
 XX
 KW Amylosucrase; EC 2.4.1.4; fusion protein; GST; glutathione-S-transferase;
 KW poly(1,4-alpha-glucan); film production; food additive; cyclodextrin.
 OS Unidentified.
 XX
 PN WO200125449-A2.
 XX
 PD 12-APR-2001.
 XX
 PF 04-OCT-2000; 2000WO-EP09695.
 XX
 PR 07-OCT-1999; 99DE-1048408.
 XX
 PA (AXIV-) AXIVA GMBH.
 XX
 PI Bengs H, Polakowski T, Held A, Gallert K;
 XX
 DR WPI; 2001-328330/34.
 DR N-PSDB; AAF61712.
 XX
 PT Amylosucrase immobilized as fusion protein with anchoring sequence,
 PT useful in continuous preparation of poly(1,4-alpha-glucan) from sucrose
 XX
 PS Claim 2; Page 34-38; 38pp; German.
 XX
 CC This invention describes a novel amylosucrase (AS), immobilizable on a
 CC solid phase, which comprises a fusion protein (FP) of functional units
 CC of AS, an anchoring sequence, and optionally additional auxiliary
 CC sequences. The invention also describes (1) nucleic acid (I) encoding
 CC FP; (2) expression vector containing (I) and able to express FP in a
 CC host cell; Escherichia coli containing the vector of (2); (3) anchoring
 CC sequence, or its functional variants or fragments, of at least 8
 CC nucleotides that encodes an epitope, a high-affinity binding partner or
 CC GST (glutathione-S-transferase); (4) solid phase for immobilizing AS
 CC comprising glutathione-Sepharose; (5) combined, stable catalyst (A)
 CC comprising FP immobilized on Sepharose for production of
 CC poly(1,4-alpha-glucan) (II); and (6) biocatalytic production of (II).
 CC AS is used in production of poly(1,4-alpha-glucans) (II), useful for
 CC producing films, as food additives, as starting materials for
 CC cyclodextrins and as auxiliaries in pharmaceutical formulations.
 CC Immobilized AS makes possible efficient, inexpensive and continuous
 CC production of poly(1,4-alpha-glucans) (II), and it can be used
 CC repeatedly. Compared with known methods, specificity is improved
 CC (increased yield of (II) and reduced formation of patainose) and
 CC reaction is complete within 24 hours, compared to 48-72 hours for batch
 CC methods. This sequence represents the expression vector construct
 CC pGEX-4T-1-Amsu5 which contains the amylosucrase (EC 2.4.1.4) described in
 CC the invention.
 XX
 XX Sequence 856 AA;
 XX
 Query Match 27.8%; Score 1208.5; DB 22; Length 856;
 Best Local Similarity 39.0%; Pred. NO. 1.5e-82;
 Matches 331; Conservative 75; Mismatches 239; Indels 203; Gaps 29;
 XX 5 PILGYWKIGLVQPTLLLEYLEEKYEHHYERDEGDKRNKKFELGFPNLPYYIDGD 64
 DB 3 pilgywkiglvqptllleyleekyeehhyerdegdkwrnkkgfelpalpyidgd 62
 XX
 QY 65 VKTQSMALIRYIADKHMLGGCPKRAEISMLEGAVLIDTRYGVSRAYSKDFETLKVD 124
 DB
 QY

Db 63 vkltqsmaliiryiadkhmlggcpkeraeismlegavldirvgvsriayskdfetlkvd 122
 QY 125 LSKLPEMLKMFEDRLCHKTYINGDHVTHPDPMILYDALDVLVYMDPCLDAFPKLVCFKKR 184
 Db 123 lsklpemlkmfeurichktylmgdhvthpdpmilgdalvlymdpmcldafpklvcfkk 182
 QY 185 IEATPIDKYLKSSKYIAMPLOQWQATFGGDDHPPKSDHILVPRGS-----MLRLKKENI-- 238
 Db 183 iealpidkylksskylawplggwqatfggdhpkksd-lvprgsbefmllqyikrild 241
 QY 239 -----QTSKRIVFTILKQKAFKGNHPCVPSVCTITYSRFHCLPDTLKS 281
 Db 242 iypqragiekedwrgfstrm-----dthfplmmeldsvygnneallpmlm 291
 QY 282 LLLPMSSKTTLSMLPOVNIIGANSPSAETPDLKKENETE--LAN--ISGPHKKSTSTSTRK 337
 Db 292 llaqawqs---ysqrn-----sslkddiatenpndwilsnkqvggcyvdlfagdik 341
 QY 338 RARSSKKKATDSVSDKT---DESVASYDSSSTHLRRSRKKPKVNYNSSSESE----- 386
 Db 342 -----gldkklpyfqlgltylhlmplfkpegksdggavssyrdvnpalgti 390
 QY 387 -----SEQISKATKKVKQKEEYVEEDVDEKSLKNSSSDEFEPVPEQLETP 436
 Db 391 gdlreviaalheagisavvdfifnhtsneh--ewagrcaaagdpfdnfyvifpdr-rmpd 447
 QY 437 SKRRSRSSAKNLEKSTMMIDHAPREMFDCLDKPIPWGRGLGYACLN----- 486
 Db 448 qydrtlreifpdghpggfsqled-----grwvwttnsfqwdlnysnp 490
 QY 487 -ILRSMKRVFCSTCRITTIQRDGLSEV-KQLGT--QNVLDLILKIVE-WNNHFGIHEMR 541
 Db 491 wvframagemlfanlgvdlrmdavafkwqmgtsenlpqahalirafn-----avmr 545
 QY 542 VSSDLFPFASHA-----KGY-----TLEFAQSHLEEVCKL--ANKYNH 578
 Db 546 iaapavfksealvhpqvvgyigqdcqigynplgmallwntlatrevnlhqltyrh 605
 QY 579 RLTHWPGQYQIASFRVWVDSALRLDAY-----HDEILSRMKLNEQLNKDAVLIIHL 631
 Db 606 nlpehtawvnyvrshddigwtfdadedaaaylgisgydhrqfnnrfvr----- 653
 QY 632 GGTEGKKETLDRFRKNYQ-----RLSDSVKARLVLENDVSWSDQLLLPLCQELNPLVL 687
 Db 654 ---fdgsfargvpfgyqpnstgdcvsgtaaalvglaqddp--havdrkilysl----- 702
 QY 688 DWHHINIVPGTLREGSLDLMLP---IPTIRE-----TWTRKGITOKOHYSES 731
 Db 703 -----alstggipilygdevglindddwsqdsksddsrwahrpnynealyaq 752
 QY 732 ADPTAISG 739
 Db 753 ndpstaag 760
 XX
 AC AAR79948;
 XX
 DT 26-FEB-1996 (first entry)
 XX
 DE GL7 antigen.
 XX
 KW Syphilis; GL5 antigen; membrane antigen; glutathione-S-transferase;
 KW GST; fusion protein; membrane antigen; diagnosis.
 XX
 OS Chimeric Treponema pallidum;
 OS Chimeric unidentified bacterium.
 XX
 FH Key Location/Qualifiers
 FT Protein 1..228

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:06:09 ; Search time 134.53 seconds
(without alignments)
551.283 Million cell updates/sec

Title: US-09-724-296-2_COPY_230_828

Perfect score: 3112

Sequence: 1 MLRLKRNQISKRIVFTIL.....KRLTARKRRSRKEVEDEK 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_Organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3112	100.0	599	3 Q10988	Q10988 schizosacch
2	1011.5	32.5	656	3 Q01408	Q01408 neurospora
3	359	11.5	325	2 Q9L4F0	Q9L4F0 bacillus ce
4	359	11.5	325	2 Q9L4E9	Q9L4E9 bacillus ce
5	306.5	9.8	322	2 Q9K9P8	Q9K9P8 bacillus ha
6	229.5	7.4	305	2 Q9S0M9	Q9S0M9 deinococcus
7	229.5	7.4	326	2 Q9RTE6	Q9RTE6 deinococcus
8	164	5.3	1588	11 Q9ESK9	Q9ESK9 mus musculu
9	156.5	5.0	368	5 Q45198	Q45198 caenorhabdi
10	155	5.0	1641	5 Q9GRZ3	Q9GRZ3 caenorhabdi
11	154.5	5.0	782	5 Q26104	Q26104 plasmodium
12	149.5	4.8	782	5 Q9U414	Q9U414 plasmodium
13	148.5	4.8	911	3 Q06704	Q06704 saccharomyc
14	148.5	4.8	1359	5 Q02061	Q02061 caenorhabdi
15	148	4.8	1359	5 Q9U7E0	Q9U7E0 caenorhabdi
16	147.5	4.7	782	5 Q26007	Q26007 plasmodium
17	145.5	4.7	401	3 Q13885	Q13885 schizosacch
18	145.5	4.7	782	5 Q9U430	Q9U430 plasmodium
19	145	4.7	1957	5 Q04010	Q04010 onchocerca

20	144.5	4.6	782	5 Q25730	Q25730 plasmodium
21	144.5	4.6	782	5 Q25875	Q25875 plasmodium
22	144.5	4.6	816	5 Q26026	Q26026 plasmodium
23	144	4.6	846	4 Q75130	Q75130 homo sapien
24	141.5	4.5	826	13 Q9YHD5	Q9YHD5 rana catesb
25	141.5	4.5	1330	5 Q22342	Q22342 caenorhabdi
26	141	4	653	2 Q34097	Q34097 streptococc
27	141	4.5	1197	3 Q9C3V7	Q9C3V7 candida alb
28	141	4.5	1205	12 Q69513	Q69513 human herpe
29	141	4.5	1702	11 Q54875	Q54875 rattus norv
30	141	4.5	4833	11 Q9QYX6	Q9QYX6 mus musculu
31	141	4.5	5038	11 Q9QYX7	Q9QYX7 mus musculu
32	140.5	4.5	1463	5 Q9GYZ0	Q9GYZ0 strongyloce
33	140.5	4.5	1680	5 Q9U8G1	Q9U8G1 plasmodium
34	140.5	4.5	2441	5 Q96124	Q96124 plasmodium
35	140	4.5	399	5 Q9VG84	Q9VG84 drosophila
36	139.5	4.5	635	10 Q40363	Q40363 medicago sa
37	139.5	4.5	643	5 Q9VCP0	Q9VCP0 drosophila
38	139	4.5	1621	5 Q01900	Q01900 caenorhabdi
39	138.5	4.5	782	5 Q9U429	Q9U429 plasmodium
40	138	4.4	1205	12 Q56300	Q56300 human herpe
41	137.5	4.4	480	13 Q91749	Q91749 xenopus lae
42	137.5	4.4	610	10 Q9LIH8	Q9LIH8 arabidopsis
43	137.5	4.4	782	5 Q9U431	Q9U431 plasmodium
44	137.5	4.4	1151	4 Q9HCL1	Q9HCL1 homo sapien
45	137.5	4.4	1416	4 Q9BZF9	Q9BZF9 homo sapien

ALIGNMENTS

RESULT	1
Q10988	AC Q10988; P87339; PRELIMINARY; PRT: 599 AA.
DT	01-NOV-1996 (Tremblrel. 01, Created)
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT	01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE	UV-ENDONUCLEASE.
GN	UVDE OR UVE1+.
OS	Schizosaccharomyces pombe (Fission yeast).
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;
OC	Schizosaccharomycetes.
OX	NCBI_TaxID=4896;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Sp972;
RX	MEDLINE=96188860; PubMed=8614629;
RA	Takao M., Yonemasu R., Yamamoto K., Yasui A.;
RT	*Characterization of a UV endonuclease gene from the fission yeast
RT	Schizosaccharomyces pombe and its bacterial homolog.;
RL	Nucleic Acids Res. 24:1267-1271(1996).
RN	[2]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=97175806; PubMed=9023111;
RA	Davey S., Nasse M.L., Ferrer J.V., Sidik K., Eisenberger A.,
RA	Mitchell D.L., Freyer G.A.;
RT	*The fission yeast UVDR DNA repair pathway is inducible.;
RL	Nucleic Acids Res. 25:1002-1008(1997).
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=972H-;
RA	Lyne M., Rajandream M.A., Barrell B.G., Lucas M., Gaillardin C.;
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR	EMBL: D78571; BAA11415.1; -
DR	EMBL: U78487; AAC49664.1; -
DR	EMBL: AL023859; CAA19577.1; -
KW	Endonuclease.
SQ	SEQUENCE 599 AA; 68816 MW; C09ED7219D5895BE CRC64;

Query Match 100.0%; Score 3112; DB 3; Length 599;

Db 276 KPFLIIIAKKNHNFDMIESKQKDIALFQL 305

RESULT 5

Q9K9P8 PRELIMINARY; PRT; 322 AA.

AC Q9K9P8;

DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (11:MBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)

DE BH2597 PROTEIN.

GN BH2597.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=86665;

RN [1]

RP SEQUENCE FROM N.A. 9153;

RC STRAIN=C-125 / JCM 9153;

RX MEDLINE=20512582; PubMed=1058132;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

RA Fuji P., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.;

RT *Complete genome sequence of the alkaliphilic bacterium Bacillus

RT halodurans and genomic sequence comparison with Bacillus subtilis.;

RL Nucleic Acids Res. 28:4317-4331(2000).

DR EMBL: AP001516; BAH06316.1. -.

KW Complete proteome.

SQ SEQUENCE 322 AA; 36900 MW; 4B0DC9291CD1E7B4 CRC64;

Query Match 9.8%; Score 306.5; DB 2; Length 322;

Best Local Similarity 26.6%; Pred. No. 5.3e-12;

Matches 85; Conservative 75; Mismatches 123; Indels 37; Gaps

Qy 247 RGRIGYACLTILSRMKERVFCSTCRITTIQR-----DGLSVKQLGTGNVLDLKLVE 301

Db 2 RIQPGYVAMSMELANASP---SKTWTATQFEKIEDHEAGRLKLERIAKTNLHNCRLRLK 57

Qy 302 WNNFGIHFNRVSSDLFPFASH---AKYGYTLEFAQSHLEEVCKLANKYNHRLTMHPGOY 358

Db 58 HNLAYQISFFRLSKSLPVLVNHPLTEGWKYELATAE-ELQAVGEFASEHOMRIDFHPDIHF 116

Qy 359 TQTASPREVVVDSAIRDLAYHDEILSRMKLEQNLKDAVLIIHLGGTFEGKKEKTLDRFRK 418

Db 117 VVLNSEAKEITRRSLQTLHYHKLKGMEDPRHR---CVLHVGGKKKGVEAGLEQFIE 172

Qy 419 NYQRLSDSVKARVLENDVSVSWSDQLPLLCQELNPLVLDMHHHNIIVPQTLREGSLDLM 478

Db 173 NTASIPKSLLSMILENDKDSYTIIDVLYLGEKLAIPVLDIIHHIDL-----HRSKSLQ 227

Qy 479 PLIFTIRETWTKGITOKEYSESADPTATSGMKRAH-----SDRVFDF-----PPCDP 528

Db 228 ETWQRTVATWDSPLPKYIHLSD-----SPISGEDDPRHHDYINADRTAFLEHIGDAVD 282

Qy 529 TMDLMIEAKEQEAVFELCR 548

Db 283 HLHVWIEAKKDLALFOLMK 302

RESULT 6

Q9S0M9 PRELIMINARY; PRT; 305 AA.

AC Q9S0M9;

DT 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)

DE UV-ENONUCLEASE.

GN UVSCDE.

OS Deinococcus radiodurans.

OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.

OX NCBI_TaxID=1299;

RN [1]


```

Db 947 ELAAKRRRAEAAA... 1004
Qy 196 EPVPEQLTEP-----ISKRRSRSSAKNLEKSTM----- 226
Db 1005 IFVEPSTLEKPERNGTTEERVMTADQAFMDAKIEBARRSRMSTRDCSVVSTLGPVKSK 1064
Qy 227 ---NLDD--HAPREMFCDLKPWPGRGL-----GYACLN 256
Db 1065 ASQRLHDIIEGKEELEDSDMDDPTNNGTLAGILYPMRSERAEVSSNHRSEGAGGMSLK 1124
Qy 257 TILSRMKERVFCSCTRTITTTQRDGLSVKQLGTONVL---DLIKLVE-----WNNHFG 307
Db 1125 HHLARKNELKEANVARSEILKAVVRORREIGVPTTILMSKSAIELVEDEDEKERNKHN 1184
Qy 308 -----IHPNRVSSDLFPFASIAKGYTLEFAQSILE-----EVGKLANKYNH---RLTMH 354
Db 1185 KTLSPHDYVRSKNEAEKAEFGK--GGTMRITNRNLKMLTRQFDLPKMSRFRKPVRIIRH 1243
Qy 355 PG-----QYTOI---ASPREV-----VVDSATRLDAYH 379
Db 1244 PNGMATIISCYDYNQIKQHLGNEMKHFPERQVRLGFAENNGVPLFAIGVMENAAE--ALH 1301
Qy 380 DEI-----LSRMLNEQLNKDAVLIILHGTFEGKKTLDLRFKRYNQLRSDSVKARL 431
Db 1302 DQFELAKNSPNTQVKGSLTNKQFIETPMKYYESAMETLDMGTFRFGPLMS--LSMV 1359
Qy 432 VLENDVSVSQDLPICQELN----IPLVLDWHHHNIVPCTLRGSLDMLPLIPIRET 487
Db 1360 GTKNEEAGNFKEML---DALNAAPFGTIPMGWDFSEVQG--IKEDTSDGQIP----- 1409
Qy 488 WTRKG-----ITOKHYSSESADPTAISGMKRR-----AHSDFRDFPPCDPT 529
Db 1410 WVRPGQMVPTDGNRSTPRPLATGRNDRRETAFTNDRINAHADQVRESTEDDPT 1465

```

```

RESULT 11
ID Q26104 PRELIMINARY; PRT: 782 AA.
AC Q26104;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE RHOPTRY ASSOCIATED PROTEIN 1.
GN RAP-1.
OS Plasmodium reichenowi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5854;
RN [1]
RP SEQUENCE FROM N.A.
RA Shi Y., Lal A.A.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20986; AAA63682.1; -.
SQ SEQUENCE 782 AA; 90160 MW; DE6D1BE2FAC308A9 CRC64;

Query Match 5.0%; Score 154.5; DB 5; Length 782;
Best Local Similarity 20.9%; Pred. No. 0.074;
Matches 147; Conservative 94; Mismatches 240; Indels 221; Gaps 37;

```

```

Qy 43 FHCLPDTLKLSPKSSKTTLSMLPQVNIIGANSFSAETPVDLKKENETELANISGPHKST 102
Db 122 FKASPSVVKTTPTGTHTS-----GSKSSS---PSSTKSSPSNVKTAS--PHGESN 168
Qy 103 STSTRKARRSKKATDSVSDKIDSVAS-----YDSSTHLRSSRS----- 144
Db 169 S--SEESTSKSRASVSGIIGVGADEEVPAPKNLTLPLEELYPTNVNLFNYKYSLNME 227
Qy 145 -----KKPNYNSSSESESEQISKATKKVKQKEEEVEEVEDEKSLNE- 189
Db 228 NINILKNGQLVAQKEEFDYDENMEKAKEER-QKALEKIGETDEEPPMTDDFLFNQV 286
Qy 190 -----SSSDEFEPVP-----EOLE-TPISKRRRSRSAAKLEKSTM-----NLDHHA 232

```

```

Db 287 KERNVAGSFRFSKLNLPFKKDEVTEKTEVSRKTYSGTGFVNDREAKILGVGATYQETP 346
Qy 233 PREMFDCLQKPIPWGRGLGYACLNILSRMKERV--PCSRFTCRITTIQRDGLSVKQLGT 290
Db 347 ETMLNCPNN-----SHLFDITSSLOGKLDIKKRENMISTTFEQOQEKLNMG- 395
Qy 291 QNVLDL-IKLVENH-----NFGIHFMRV---SSDLFPFASHAKY-----GYTLE---F 332
Db 396 --VLDLELADTECKFGTCIGSRGEHHLRLYEFDLFPHPNIDYLTLDAGYKQKNNHIY 453
Qy 333 AQSH-----LEEVGKLAN-----KYNHR-----LTMHPGQ--YTIQI 361
Db 454 ELSHVNFCLLNPKTLEEFLLKKEIKDLMGDDLLIKYKFNDFNPMFISITCHIESLIYDDI 513
Qy 362 ASPREV-----VDSAI-----RDLAYH-----DELSRMKL---NEQLN 393
Db 514 EASQDVAALVLIKAKSKLHVITSGLSYKARKLVYKIYSEIQKNPDDELYEKLWIYDIYML 573
Qy 394 KDVAVLIHGGGT---PEGKKETLDRFRKNYORLSDSVK-----ARLVLEN---DDVSWS 441
Db 574 KRYIYAYALEGVCSYLENDKTMQMDTELHIYNKLVDSVRYSSCFKNVIYVNAIMSGIHEK 633
Qy 442 VQDLPLQCELNIPVLVDWHHHNIVPCTLRGSLDMLPLIPIRETWTGKITQKHYS 501
Db 634 IKHFTKLVPRQN--FLLDYHFNISF-----EKEIKPVKKYST 668
Qy 502 S---ADPTAIS---GMKBRASHDRVDFPPCDPTMDLMEAKEKEQAVFELCRRYELON 554
Db 669 SHVYDPTVAVSYAYNLDRTWTVTIINDY-----FEAKKRELIVIVSRMKTDM-- 716
Qy 555 PCPCEIMGPEYDQTRDGYYPGCAEKRLTARKRSRKEEVEE 596
Db 717 ----LSLQNDSEKIPND----KSANSKLATRLMRKFAEIRD 750

```

```

RESULT 12
ID Q9U414 PRELIMINARY; PRT: 782 AA.
AC Q9U414;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE RHOPTRY ASSOCIATED PROTEIN-1.
GN RAP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Li X.R., Luo S.H., Yu X.B., Shan Z.X., Ma C.L.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF206631; AAF15365.1; -.
SQ SEQUENCE 782 AA; 90082 MW; 8EIF4CF2883903FD CRC64;

Query Match 4.8%; Score 149.5; DB 5; Length 782;
Best Local Similarity 21.0%; Pred. No. 0.15;
Matches 147; Conservative 94; Mismatches 241; Indels 219; Gaps 39;

```

```

Qy 43 FHCLPDTLKLSPKSSKTTLSMLPQVNIIGANSFSAETPVDLKKENETELANISGPHKST 102
Db 122 FKASPSVVKTTPTGTHTS-----GSKSSS---PSSTKSSPSNVKTAS--PHGESN 168
Qy 103 STSTRKARRSKKATDSVSDKIDSVS---ASYDSSTHLRSSRSRKKPVNYN---SSSES 156
Db 169 S--SEESTSKSRASVSGIIGVGADEEAPPAPKNLTLP-----EELYPTNVNLFNYKYSLN 223
Qy 157 ESEEQISKATKK---VKQKEEEVEEVD-----EKSLLKNSSSDSEPEPVPEQ---LE 204
Db 224 NMEENITLKNEGDLVAQKEEPEYDENMEKAKODKKALEKTGKOSDEEPPMFSENKFELE 283

```



```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Favello A., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierri-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX STRAIN=BRISTOL N2;
RA Fulton R., Wohlmann P.;
RT "The sequence of C. elegans cosmid B0041.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO HELICASE C-TERMINAL DOMAIN.
DR EMBL; AF000196; AAC24256.1; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
KW ATP-binding; Helicase.
SQ SEQUENCE 1359 AA; 156235 MW; 5F2ACE099EDED444 CRC64;

Query Match 4.8%; Score 148.5; DB 5; Length 1359;
Best Local Similarity 20.9%; Pred. No. 0.36;
Matches 136; Conservative 96; Mismatches 222; Indels 197; Gaps 34;

QY 83 LKKNETELANTISGPHKKSTSTSTRKRARSSKKKATDSVSDKIDSVASVDSSTHLRSS 142
DB 229 VKKESESE---DEAEKK--KTEKKRKRKTSSESESE--SEKSDSEEEKESSP-----KP 277

QY 143 RSKKPVNYS--SESESEEQ-----ISKATKKVKOKEEVEEVEEDVDEKS 185
DB 278 KKKKPLAVKKLSDESEESDVEVLPOKKRGAVTLISDSEDKDKGSESE--ASDVEKV 336

QY 186 LKNESSEDEFEPVPEQLTPTISKRRSRSSAKNLEKSTWNLD-----DHAPR 234
DB 337 SKKAKKQSSSGSSGSSGIVNRRKSKKKKPEKKRKGIIIMDSKLOKETIDAEK 396

QY 235 EMFDCIDKIPWRGLGYACLTNTI-----LRSKRVFCRSTRTITTIQDGL----- 283
DB 397 ERRKRLEKK-----QKEFNGVILEGEDLTMLTGTSSQKLSVLPDPSDSTVEE 448

QY 284 SVKQLGTQNVLDLIKLVENHNFHFRVSSDLFPFASHAKYGYTFEFAQ----- 334
DB 449 SKKPEVHN--SLVRILKPHQAHGQIFM-----YDCAFESLDRLDTGEGS 490

QY 335 -----SHLEEVGKLAN--KYNHRLTWRP--GOYTQ---IASPREVVVD---SAIRLAYHD 380

```


Db 259 SESEKSDDEEEE 270

Search completed: January 15, 2002, 14:06:13
Job time: 952 sec



Db 241 DKPIPWGRGLYACVNTILRSKMERVFCSTRTCTTTIQRDGLSVKQOLGTONVLDLKLK 300
QY 301 EWNHNFGEHMRVSSDLPPFASHAKYGTLEFAOHLSEVGLKANKYNHRLTMHPGQYTO 360
Db 301 EWNHNFGEHMRVSSDLPPFASHAKYGTLEFAOHLSEVGLKANKYNHRLTMHPGQYTO 360
QY 361 IASPREVVDSAIRDLAYHDEILSRMKLNEOLNKDAVLIHLGGTFEGKKTLDLRFKKN 420
Db 361 IASPREVVDSAIRDLAYHDEILSRMKLNEOLNKDAVLIHLGGTFEGKKTLDLRFKKN 420
QY 421 QRLSDSVKARLVLENDVSWSVQDLPCLQELNIPVLVDWHHNTVPGTLREGSLDLMPL 480
Db 421 QRLSDSVKARLVLENDVSWSVQDLPCLQELNIPVLVDWHHNTVPGTLREGSLDLMPL 480
QY 481 IPTIRETWKGIITOKOHSYSESADPTAISGMKRRRAHSRDFPCCDPTMDLMEAKEKE 540
Db 481 IPTIRETWKGIITOKOHSYSESADPTAISGMKRRRAHSRDFPCCDPTMDLMEAKEKE 540
QY 541 QAVFELCRRYELQNPCCPLEIMGPYDQTRDGYPPGAEKRLTARKRRSRKKEEVEEDEK 599
Db 541 QAVFELCRRYELQNPCCPLEIMGPYDQTRDGYPPGAEKRLTARKRRSRKKEEVEEDEK 599
RESULT 2
S55262
UV-endonuclease - Neurospora crassa
C:Species: Neurospora crassa
C>Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
C:Accession: S55262
R:Yajima, H.; Takao, M.; Yasuhira, S.; Zhao, J.H.; Ishii, C.; Inoue, H.; Yasui, A.
EMBO J. 14, 2393-2399, 1995
A:Title: A eukaryotic gene encoding an endonuclease that specifically repairs DNA damage
A:Reference number: S55262; MUID:95292980
A:Accession: S55262
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-656 <YAJ>
A:Cross-references: EMBL:D11392; NID:g1526560; PID:g927215
C:Genetics:
A:Gene: uvel

Query Match 32.5%; Score 1011.5; DB 2; Length 656;
Best Local Similarity 36.6%; Pred. No. 3.2e-48;
Matches 241; Conservative 88; Mismatches 149; Indels 181; Gaps 18;

QY 110 ARSKKKATDSVSKIDESVASYDSST-----HLRPSR-----SKKPVNYNSSSESP- 157
Db 3 SRKSAALDTPSESTSTSSLDSSAPSPARNLRSGRNILQPSSEKDRDHEKRSGBEL 62
QY 158 -----SEQISKATKKVKQKEEVEEVEEVEDEK 184
Db 63 AGRMGKANGCHLREBKEQEGVKMAIEGLARMERLQRAKQKQLEED----- 114
QY 185 SLKNSSDEFPVPEQLETPISKRRS-----RSSANKLEKSTMNLDHIA----- 232
Db 115 GIPVPSVSRF-PTAPYHKKSTNAEERAKEPVLKTHSKDVEREAIEGVDVVKVMEPAAT 173
QY 233 -----PREMFCDLCK-----PIPWGRGLYACVNTILRSKMERVFCSTRTCTIT 275
Db 174 NIEPEDAQDAERGAARPPAVNSVYLPWKGRLGYACVNTILRSKPPFSRTRCMA 233
QY 276 TI-----QRDGLSVKQOLGTONVLDLKLIVVHNNHNG 307
Db 234 SIVDHRHPLQFEDEPEHLKKNPKDSKEPODELGHKFVQELGLANARDIVKMLCNNEKYG 293
QY 308 IHMRVSSDLPPFASHAKYGTLEFAOHLSEVGLKANKYNHRLTMHPGQYTOIASPRE 366
Db 294 IRFURLSEMPPFASHVGHYKLAPASEVLAEGRAVAAELGHRLLTHPGQFTOLGSPRK 353
QY 367 VVDSAIRDLAYHDEILSRMKLNEOLNKDAVLIHLGGTFEGKKTLDLRFKKNYQRLSDS 426
Db 354 EVVESAIRDLEYHDEILSLKLPOQNRAVMIHMGQFGDKAATLRFKKNYARLSQS 413

QY 427 VKARLVLENDVSWSVQDLPCLQELNIPVLVDWHHNTV--PGTLREGSLDM--PLIP 482
Db 414 CKNRVLVLENDVQWTVHDLPLVCEELNIPVLDYHHNICFDPFAHLREGTLDISDPKLOE 473
QY 483 TIRETWKGIITOKOHSYSESADPTAISGMKRRRAHSRDFPCCDPTMDLMEAKEKEQA 542
Db 474 RIANTKRGKIGKMHYSEPCD--GAVTPRHRRHRPRVMTLPPCPPDMDLMEAKEKEQA 532
QY 543 VFELCRRYEL-----QNPCCP-----LEIMGPE--- 565
Db 533 VFELMTFTKLPGEKINDMVDPYRDDENRAPPVKKAPKKKGGKRRKRTTDEAAAEVVD 592
QY 566 --YDQTRDG-----YPPGAEKRLTARKRRSRK-----EEVEED 597
Db 593 TAADDVKDAPEGKPEVEERAMGPNRVYWPGLGCEEWLKPKKREVKKGVPEVEDE 651
RESULT 3
S55418
UV-endonuclease homolog ywjd - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 15-Oct-1999
C:Accession: S55418; B70060
R:Glaser, P.; Danchin, A.
submitted to the EMBL data Library, May 1995
A:Description: Cloning and sequencing of the Bacillus subtilis chromosomal region fro
A:Reference number: S55414
A:Accession: S55418
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-320 <GLA>
A:Cross-references: EMBL:249782; NID:g853752; PID:CAA89865.1; PID:g853757
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabbet, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koettel, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:98044033
A:Accession: B70060
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-320 <KUN>
A:Cross-references: GB:299123; GB:AL009126; NID:g2636240; PID:CAB15748.1; PID:ell862
A:Experimental source: strain 168
C:Genetics:
A:Gene: ywjd

Query Match 11.6%; Score 360; DB 2; Length 320;
Best Local Similarity 29.8%; Pred. No. 6.6e-13;
Matches 84; Conservative 61; Mismatches 109; Indels 28; Gaps 8;

QY 276 TIQRDGLSVKQOLGTONVLDLKLIVVHNNHNGIHFMRVSSDLPPFASH--AKYGYTLEFA 333
Db 34 TERKALLTVTKANLRNTM---RTLHYIIGHGIPLYRESSIVPLATHPDVWMDVETPF- 89
QY 334 QSHLEEVGKLANKNYHRLTMHPGQYTOIASPREVVDSAIRDLAYHDEILSRMKLNEQLN 393
Db 90 QKEFPEIGELVKTHQRLTSEHPNQFTLFTSPKESVTKNAVTDMAHYRMLEAMGIADR-- 147
QY 394 KDAVLIHLGGTFEGKKTLDLRFKKNYQRLSDSVKARLVLENDVSWSVQDLPCLQELN 453
Db 394 KDAVLIHLGGTFEGKKTLDLRFKKNYQRLSDSVKARLVLENDVSWSVQDLPCLQELN 453

Db 148 --SVINIHIGAYGNKDTATQAQFONIKOLPOEIKERMTLENDKTKVTYTTETLQVCEQED 205
 QY 454 IPLVLDWHHHHNVPTLREGSLDLMLPLIPTIRETWTRKGITOKOHYESADPTAISGMKR 513
 Db 206 VPFVDFHHFYANP-----DHADLNLVAPRMKRTWIRIGLQPKVHLSPKSEQAI----- 256
 QY 514 RAHSDRV---PDDFPD-----PTMDLMEAKKEQAVFEL 546
 Db 257 RSHADYVDANFLPLERFRONGTNIEMTEAKOKKALLRL 298

RESULT 4
 E83374
 hypothetical protein BH2597 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
 C:Accession: E83974
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650; MUID:20263314
 A:Accession: E83974
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-322 <STO>
 A:Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BA806316.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH2597

Query Match 9.8%; Score 306.5; DB 2; Length 322;
 Best Local Similarity 26.6%; Pred. No. 5.7e-10;
 Matches 85; Conservative 75; Mismatches 123; Indels 37; Gaps 9;

QY 247 RGRIGYACLTILRSKMERVFCSTCRITTIQIORDGLESVKOLGTQNVLDLIKLE 301
 Db 2 RIQFGYVAMSMELANASP---SKMTATQFEKIEDHEAGRLKRLERIAKTNLHCLRLK 57
 QY 302 WNNHFGTHFRVSSDLPFPASH---AKGYTLEFAOSHLEEVCKLANKYNHRLTMIPGQY 358
 Db 58 HNLAYQISFRLESKLVPLVNHPLTEGKRYELATAE-ELQAVGEFASEHQMRIDFHPDH 116
 QY 359 TQTASPREVVDSAIRDLAYHDEILSRMKLNEQLNKDAVLIHLGCTFECKKETLDRFRK 418
 Db 117 VLNSEAKETRRSLQTLVHYKLLGMEIDPRHR---CVLHVGGKKGVAGLEQFIE 172
 QY 419 NYORLSDSVKARLVLENDVSVQDILLPCQELNIPVLVDWHHHNIVPGTLREGSLDLM 478
 Db 173 NTASIPKSLSMINLENDKSYTIDVLYLGEKLAIPVVDLIIHHDLV-----HRSKSLQ 227
 QY 479 PLIPTIRETWTRKGITOKOHYESADPTAISGMKRRAH-----SDRVDF-----PPCDP 528
 Db 228 ETWQRIWATWEDSPLPKVHLS-----SPLSGEDDPRHHDYINADRFIAFLHEIGADAVD 282
 QY 529 TMDLMEAKEKEQAVFELCR 548
 Db 283 HLHVMIEAKKKDLALFOLMK 302

RESULT 5
 C73350
 probable UV damage endonuclease - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
 C:Accession: C73350
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896
 A:Accession: C73350

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-326 <WHI>
 A:Cross-references: GB:AE002022; GB:AE000513; NID:g6459590; PIDN:AAE11370.1; PID:g645
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR1819
 A:Map position: 1

Query Match 7.4%; Score 229.5; DB 2; Length 326;
 Best Local Similarity 27.9%; Pred. No. 9.6e-06;
 Matches 88; Conservative 42; Mismatches 140; Indels 45; Gaps 10;

QY 249 RLGYACLTILRSKMERVFCSTCRITTIQIORDGLESVKOLGTQNVLDLIKLEWHHHNFI 308
 Db 31 QLGLVCLTVGPEVRFTVTLRSRYALSPAAREA--KLDDLYSSNIKTLRGAADYCAAHDI 88
 QY 309 HPMRVSSDLFP---FASHAKYGYTLEFAOSHLEEVCKLANKYNHRLTMHPGOYTQIASPR 365
 Db 89 RLYRJSLSLFPMLDLAGDDTGAAVLTHLAPQLLEAGHAFTDAGVRLLMHPEQFIVLNSDR 148
 QY 366 EVVDSAIRDLAYHDEI-----LSRMKLNELNKDAVLIHLGCTFECKKETLDRFRKNY 420
 Db 149 PEVRESSVRAMSAHARVMDGLGARTPN-----LLHCGKGGCGAE-----LAALI 196
 QY 421 QRLSDSVKARLVLENDVSVQDILLPCQELNIPVLVDWHHHNIVPGTLREGSLDLM- 479
 Db 197 PDLPPDVRRLGLENDERAYSPAELLPTCEATGTPLVFDAAHHVV-----HDKLPD 247
 QY 480 -LIPTIRETWTRKGIT-----OKOHYESADPTAISGMKRRAHSDRVDFPPCDPTMD- 531
 Db 248 QEDPSVVRWLRARATWQPEWQVHLS-----NGIEGPQDRRHSHLIADEFPAYADVPW 302
 QY 532 LMIEAKEKEQAVFEL 546
 Db 303 IEVEAKGKEEATAAL 317

RESULT 6
 S55785
 nucleolar protein gar2 - fission yeast (Schizosaccharomyces pombe)
 N:Alternate names: probable RNA-binding protein RBD18
 C:Species: Schizosaccharomyces pombe
 C:Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 17-Nov-2000
 C:Accession: S55785; S68087; T37634; S52318
 R:Gulli, M.P.; Girard, J.P.; Zabetakis, D.; Lapeyre, B.; Melese, T.; Calzergues-Ferte
 Nucleic Acids Res. 23, 1912-1918, 1995
 A:Title: gar2 is a nucleolar protein from Schizosaccharomyces pombe required for 18S
 A:Reference number: S55785; MUID:95319932
 A:Accession: S55785
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-500 <GUI>
 A:Cross-references: EMBL:Z48166; NID:g663261; PIDN:CAA88179.1; PID:g663262
 R:VanHoy, R.W.; Wise, J.A. 1996
 Curr. Genet. 29, 307-315, 1996
 A:Title: Molecular analysis of a novel Schizosaccharomyces pombe gene containing two
 A:Reference number: S68083; MUID:96171513
 A:Accession: S68087
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 374-407 <VAN>
 R:Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
 submitted to the EMBL Data Library, September 1997
 A:Reference number: T21733
 A:Accession: T37634
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-338, S', 340-500 <BRO>
 A:Cross-references: EMBL:Z99091; PIDN:CAB11772.1; GSPDB:GN00066; SPDB:SPAC13F5.09
 A:Experimental source: strain 972h-; cosmid c13F5
 C:Genetics:

A:Gene: gar2; SPDB:SPAC13F5.09

A:Map position: 1

C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein repeat-containing proteins; ribonucleoprotein repeat homology <RRM1>
F:264-331/Domain: ribonucleoprotein repeat homology <RRM1>
F:367-433/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match 4.9%; Score 154; DB 2; Length 500;
Best Local Similarity 27.9%; Pred. No. 0.23;
Matches 53; Conservative 33; Mismatches 60; Indels 44; Gaps 6;

QY 48 DTLKLLPMSKTTLSMLPQVNIANSFSAETPVDLKKENETELANISGPHKKSTSTSTR 107
DB 35 EAKEKAKOSSKTDVS--PK-----KSKKEAKRASSPEPKSV----- 71
QY 108 KRARSKKKATDSVKIDIESVASVDSSTHLRRSRKPKVNNSSSESEEEQISKATK 167
DB 72 KKOKSKKKKESSSESSSESSSESSSESSSESSSESSSESSSESSSESSSESSSESS 131
QY 168 KVKQ-----KEPEYVEEVDKSLKNFSSSDEFEPPVPEQLETPISKRRSRSS 216
DB 132 EKKESSSESSSESSSESSSESSSESSSESSSESSSESSSESSSESSSESSSESS 181
QY 217 AKNLEKSTW 226
DB 182 SESEEEVV 191

RESULT 7

T40373
probable thioredoxin - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40373
R:Wood, V.; Rajadream, M.A.; Barrell, B.G.; Hilbert, H.; Duesterhoeft, A.
submitted to the EMBL Data Library, December 1998
A:Reference number: Z21923
A:Accession: T40373

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1726 <WOO>

A:Cross-references: EMBL:295620; PIDN:CAB09115.1; GSPDB:GN00067; SPDB:SPBC3D6.13c

A:Experimental source: strain 972h; cosmid c3D6

C:Genetics:

A:Gene: SPDB:SPBC3D6.13c

A:Map position: 2

Query Match 4.8%; Score 150.5; DB 2; Length 726;
Best Local Similarity 22.3%; Pred. No. 0.59;
Matches 127; Conservative 70; Mismatches 223; Indels 149; Gaps 26;

QY 44 HCLPDT---LKS---LLPMSKTTLSMLPQVNIANSFSAETPVDLKKENETEL--ANI 94
DB 133 HLPDTPDIPSDVDVLTDEDTEVASIOPAISTSVSSLSLASTAMSKASASEPSSGV 192
QY 95 SGPHKSTSTSTRKRSKKATDSVKIDIESVASVDSSTHLRRSRKPKVNNSS 154
DB 193 TKASKKLTSSPT---SVASKKATLSSVS-----KVAS--TSSLPVTSVASVDPKS--AASK 242
QY 155 ESSEEOISKATKKVKQKEEEVEEVDKSLKNFSSSDEFEPPVPEQLETPISKRRSR 214
DB 243 VDAEFSIQAPSPFPEKEKENTETEE-----SKKSINPT-----CT 281
QY 215 SSARKLEKSTMTNLDHA-----PREMFCDLCKPWP-----RGRGLYACLNITLR 260
DB 282 SKALALDADIDAALTDKEGWFIOFYSECCDDVDSTANYANMRGKLNVAHINCAV- 340
QY 261 SMKRVFCSTCRITTTQ-----RGLSVKQIGTQNVLDLKLKLVENHGHIFM 311
DB 341 -----SKRACKQYSIQYPTFTLFTKEAF--VEYVGLPNEGDLVSFAEAAEFIR-- 389
QY 312 RVSSDLFPFASHAKY--YTFLEAQSHLEVGKLANKYNHRLMHPG-----QY 358

DB 390 --EVELLDTVNAEKNGDVFLEYDDSAEYLNIRKGTQLLGHANLYLTTSQITAKKY 447
QY 359 TQIASPREVVVDSAI-----RDLAYHDEILSRMKN-----EQLNKD 395
DB 448 RVVSFPKLIIVRDGIASYYPAKMAQDNKDYRRILGMMKNWLPVLPDELRTSNKEIFNDE 507
QY 396 AVLIHLLGGTFEGKKT-----LDR-----FRKNYQRLSDSVKARLVLENDVSW 440
DB 508 SVLFLNPELDFDETKRTAQKIATIEFLDECKTQSNWQKETDK-KNSLVNEAEKN- 565
QY 441 SVODLLPLCOELNI-----PLVLDWHHNIIVPGTLREGSLDLMLPLIPIRETWTIKGI 493
DB 566 ---DLEAIEAAKNFHVNGKDSPTRFVWNGKFAQWLKRKFIDIDIEATGPRV----IVYNA 618
QY 494 TOKOHTSESADPTAISOMKRRASDRVFD 522
DB 619 AADIYWDETAKGPIS-----IEKDVTFD 642

RESULT 8

S51441

hypothetical protein YLR309c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein L2142.5

C:Species: Saccharomyces cerevisiae

C:Date: 23-Feb-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999

C:Accession: S51441

R:Pauley, A.

submitted to the EMBL Data Library, November 1994

A:Description: The sequence of S. cerevisiae cosmid L2142.

A:Reference number: S51437

A:Accession: S51441

A:Molecule type: DNA

A:Residues: 1-911 <PAU>

A:Cross-references: EMBL:U17247; NID:g577216; PIDN:AAB67359.1; PID:g577221; MIPS:YLR3

C:Genetics:

A:Gene: SGD:IMH1

A:Cross-references: SGD:S0004300; MIPS:YLR309C

A:Map position: 12R

Query Match 4.8%; Score 148.5; DB 2; Length 911;
Best Local Similarity 20.4%; Pred. No. 1;
Matches 113; Conservative 75; Mismatches 180; Indels 185; Gaps 21;

QY 6 KRNIQISKRIVETILKOKAFKGNHPCVPVCTITYSRFHCLPDTLKSLLPMSKTTL--- 62
DB 202 KEVYSIE-----LKEBLAIKNHSLEDSRMKITELEON-----LSSKSTIMEE 244
QY 63 --SMLPQVNIANSFSAETPVDLKKENETELANISG-----PHKKSTSTSTRKR 111
DB 245 KSSSLAELNI-----TLKEKERKLSLEKMKELPKAISHONVGNRRKKNR 292
QY 112 SSKKK-----ATDSVSDK--IDESVASVDSSTHLRRSRKPKVNNSSSESESE- 163
DB 293 NKGKKNKGITTDISEETVDSINTEE-----YDKLKENLQELQEKY 336
QY 164 KATKKVKQKEEEVEEVDKSLKN---ESSDEFEPPVPEQLETPISKRRS----- 213
DB 337 KDCEDWKQYEDIEAELDKALENSQLEKSAKELETINTLEIDTKSLKKEKNESELEVR 396
QY 214 -----RSSAKNLEKEST--MNLDQ--HAPREMFCDLCKPWPWRGLG 251
DB 397 DMLRTVGNHLDVADKETEKSCKQNEEVKTVKLELDLDRHKNATIE-----A 444
QY 252 YACNLNITLRSKRVFCSTCRITTTIQDGLSVKQIGTQNVLD-----LTKLVENH 304
DB 445 YEAKNTELRSKIE-----LLSKKVEHLKNLCTEKEKEQTTSONKVAKLENEIS 492
QY 305 NFGIHFIMVSSDLDPFFASHAK-----YCYTLFEAQSHLEVGKLANKYNH 349
DB 493 QUTYEKSNITKELTSLRTSYKOKETVSYLEEQVKQFSEQKDAEKSTEQURKHAKISN 552

QY 503 ---ADPTAIS-----CMKRAHSDRVDFPPCPDPTMDLMEAKEKEQAVFELCRRYELONP 555
DB 670 HIYFOPTVASYAYNLDRTMTWIINDY-----FEAKKELTIVISRMKTD----- 716
QY 556 PCPLEIMGPEYDQTRDGYVPCGAERLRTARKRRSRKEVEE 596
DB 717 ---LSLONEESKIPND-----KSANSKLATRLMKFKAEIRD 750
RESULT 11
RN283
polymerase-associated nucleocapsid phosphoprotein - parainfluenza virus type 1 (strain C)
C:Species: Parainfluenza virus type 1
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jul-1999
C:Accession: E40234
R:Power, U.F.; Ryan, K.W.; Portner, A.
Virology 189, 340-343, 1992
A:Title: The P genes of human parainfluenza virus type 1 clinical isolates are polycistronic
A:Reference number: A40234; MUID:92295573
A:Accession: E40234
A:Molecule type: genomic RNA
A:Residues: 1-368 <POM>
A:Cross-references: GB:M74080; NID:g332679; PIDN:AAA46830.1; PID:g332681
C:Superfamily: parainfluenza virus polymerase-associated nucleocapsid phosphoprotein
C:Keywords: nucleocapsid; phosphoprotein

Query Match 4.7%; Score 147; DB 1; Length 568;
Best Local Similarity 21.2%; Pred. No. 0.66;
Matches 114; Conservative 87; Mismatches 214; Indels 122; Gaps 23;
QY 62 LSMLPQVNIAGNSAETPVDLKKENETELANISGPHKKSTSTSRKRSKKKATDSV 121
DB 36 LSKYP-TEIGEDRWLHNIIDNPENKSKSDONNKDRAISTQDH-RSEB---SGI 90
QY 122 SKIDESVASYDSSTHLR-----RSR---SKKPNVNSSE-----SSEE 160
DB 91 SRTGES-----KTTHARILQOQTHRASRGTSNPLPENMGDNRTRIDEDSPNRRH 146
QY 161 QISKATKVKQ-----KEEYVEVEDEKSLKNESSEDEFEPVPEQLETPISKRRSR 214
DB 147 QRSVPTDEDRKMAENSKREDOVEGFEVVGSTLSDDGE-----GRTN 192
QY 215 SSAKNLERESTMNLDDHAPREMFCIDKPIWRGRGLYACLTILRSKMKRVFCSTCR 274
DB 193 NNGRSMETSST-----HSTR-ITDVTNPSP-----ELEEAVALQKKRRP 231
QY 275 TTIQDGLSVKQLGTQNVLDLILKLEW-NHNFGRHMRVSSDLFPFASHAKYGYTLEFA 333
DB 232 TTIKRQTRSR---TQN-SLHKSTGSDSNLEHDHNTKTSQKIPP----- 273
QY 334 QSHLEEVGKLANKYNHRLTMHPGOYTQIASPREVVVDSAIRDLAYHDEILSRMKLNOLN 393
DB 274 SKNEEPAVTQKNHNRKTKHT---TNNANNNAKCLPTPEHDTTSNEBGTNTSDVEM-- 328
QY 394 KDAVLIHLG-----GTEGKETLDRFK-----NYQRLSDSVKARLVLENDVSWV 442
DB 329 --AKLLSVLGVMKSHQEFELSRASHOPAKRMLKSANYKENTFNLGMLLSVEKSLGNKV 386
QY 443 QDLLPLQCELNPLVLDWHHINIVPGTREGSLDMLPIPTIRETWTRKGITQKHYES 502
DB 387 EENRLLKQIOEIDSSDLRHLKRFSEYQKQNSLMMANLSTLHIITDRGGKT-----GDP 441
QY 503 ADPTAISGMKRAHSDRV-----FDFPPCPDPTMDLMEAKEKEQAVFELCRRYELONP 555
DB 442 SDTTRSPSVFTKGKONKVKKTRF-----DPSMEALGGQEFKPDILREDELREDIRNP 493

RESULT 12
T38786
translation initiation factor eif-3 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
C:Accession: T38786; T38056
R:Skellton, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z21751
A:Accession: T38786
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 280-913 <SKE>
A:Cross-references: EMBL:Z98762; PIDN:CAB11485.1; GSPDB:GN000066; SPDB:SPAC4A8.16c
A:Accession: T38056
A>Status: preliminary; translated from GB/EMBL/DDDBJ
A:Molecule type: DNA
A:Residues: 1-401 <SK2>
A:Cross-references: EMBL:Z98599; PIDN:CAB11250.2; GSPDB:GN000066; SPDB:SPAC1E11.01c
A:Experimental source: strain 972h; cosmid c1E11
C:Genetics:
A:Gene: SPDB:SPAC1E11.01c
A:Map position: 1

Query Match 4.7%; Score 145.5; DB 2; Length 918;
Best Local Similarity 23.5%; Pred. No. 1.5;
Matches 103; Conservative 57; Mismatches 172; Indels 107; Gaps 20;
QY 71 GANSFSAETPVDLKKENETELANISGPHKKSTSTSRKRSKKKATDSVSDKIDESVA 130
DB 7 GGSDDSDAESVDSSEEN-----RLTSRLKKODSSSESESESESESESESESE 61
QY 131 SYDSTHLRKRSSKKPNVNSSESESEBEQISKATKVKQKEEYVEVEDEKSLKNES 190
DB 62 SESEVEVPK---KKAASSEDSESDSE-----SSEEEETESEDSEVSESESESE 112
QY 191 SSDEFPVPEQLETPISKRRSRSSAKNLEK-----STMNLDDHAPREM 236
DB 113 ESSESESESE-ESDESERSGPSFLKPKPEKEAPAGLKLFGRESSSEDEEBGRV 171
QY 237 F-DCLDKPIWRGRGLYACLTILRSKMKRVFCSTCRITTTIQDGLSVKQLGTQNVLD 295
DB 172 VKSAKD---RYEEFISCMETIKNAMS-----SNWIVVSNFEFDHLNKVSQ----- 214
QY 296 LILKLEWNNHFG---IHFMRVSSDLFPFASHAKYGYTLEFAOSHLEEVGKLANKYNHRLT- 352
DB 215 --KCKEAGRNPPYIEFALSALDOKLESADKA-----FIKSLDAANGRAFNALKQVRK 265
QY 353 -----MHPGOYTQIASPR-----EVVWDS-AIRDLAYHDEILSRM 386
DB 266 NNFQFQSDIDRVKDPGEFMKPAELNEIPKPAKAGQDEVIVDGVATRGIVAPTEGLGK- 324
QY 387 KLNFLNKKDAVLIHLGGTTFE--GKETLDRFK-----NYQRLSDSVKARL 431
DB 325 --PEET-PADIEKYLRALFEARGKKST-DRSEQTLLEKLTIAVTDYORL--RVKVAL 378
QY 432 VLENDVSWVQDILLPLCQ 450
DB 379 LAVREDINTSGOYMPIDQ 397

RESULT 13
S48326
hypothetical protein YML071c - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 02-Dec-1994 #sequence_revision 09-Mar-1996 #text_change 29-Oct-1999
C:Accession: S48326; S48825
R:Bowman, S.; Churcher, C.
submitted to the EMBL Data Library, August 1994
A:Reference number: S48326
A:Accession: S48326
A:Molecule type: DNA
A:Residues: 1-113 <BOW>
A:Cross-references: EMBL:Z38114; NID:g558402; PIDN:CAA86249.1; PID:g558403; MIPS:YML0
R:Brown, D.; Bowman, S.
submitted to the EMBL Data Library, October 1994

A:Reference number: S48816
A:Accession: S48825
A:Molecule type: DNA
A:Residues: 88-607 <BRO>
A:Cross-references: EMBL:Z46373; NID:g587529; PIDN:CAA86507.1; PID:g587539; MIPS:YML0716
C:Genetics:
A:Map position: 13L

Query Match 4.7%; Score 145; DB 2; Length 607;
Best Local Similarity 21.6%; Pred. No. 0.93;
Matches 82; Conservative 73; Mismatches 160; Indels 64; Gaps 18;

QY 1 MLRLKLNQIS--KRIVFTILKQKAFK-NHPCVPSVCTITYSRPHCLPDL-----50
DB 258 LVKLLSTNVSNLKLKILQYLNSTPPDGKTNKSLV--FLANRYKFTIDETASYPLDV 315

QY 51 ----KSLPMSKSTLSLMPQ-----VNIGANSFSAET---PVDLKKENETELANISGPH 98
DB 316 ESSNESLIEVMVKKRIEVRHVYMSLVFLKSFLYDTNDLEIPFPELESTVLRINGTN 375

QY 99 KKSSTSTRKARSKKKATSDSKIDESVASYDSSTHLRRSSRSKKPVNYSSESSES 158
DB 376 EEKETEKEKTKKEEYKODSVANN--EEDVTENKSIDEDVOEEVOGKVE--GEDDGAERKT 433

QY 159 EEOISKAT-KVKOKEEVEEYDEKSLKNESSESSDEF--PVVPEOLETPISKRRRSRS 215
DB 434 ENETENVTNKTEKAEKEEVEVNTKDKAEKEEVEEINKEVVTPEEPSKIDKAE----490

QY 216 SAKNLEKSTNLDHAPREMFCDLCKPIPWGRGLGYACLTILRSKMERVFCRSRTCRIT 275
DB 491 -----KEEEINKVEVTPDEP-----SKKIRTSKRENKIPTNAVMLQFVDK-----C-IT 534

QY 276 TIQRD---GLSVKQLTQNVLDLI-----KLVENHNHFGIHFMRVSSDLPPFASHAKYGY 328
DB 535 YVLKDLTRGLNSIK-LSDSVCLQLYVCSFRCLDLNRYHHLFLKKNIDTSLF-----585

QY 329 TLEFAQSHLEVGKLANKY 347
DB 586 TTEQLARAIDKRAELASKY 604

RESULT 14

A59294
skeletal myosin - nematode (Onchocerca volvulus)
N:Alternate names: major body wall myosin
C:Species: Onchocerca volvulus
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000
C:Accession: A59294
R:Werner, C.; Rajan, T.V.
Mol. Biochem. Parasitol. 50, 255-260, 1992
A:Title: Comparison of the body wall myosin heavy chain sequences from Onchocerca volvulus
A:Reference number: A59294; MUID:92158005
A:Accession: A59294
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1957 <WER>
A:Cross-references: GB:W74066; NID:g159892; PIDN:AAA29420.1; PID:g159893
A:Experimental source: sex female
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:87-773/Domain: myosin motor domain homology <MMO>

Query Match 4.7%; Score 145; DB 2; Length 1957;
Best Local Similarity 18.1%; Pred. No. 4.3;
Matches 130; Conservative 111; Mismatches 231; Indels 248; Gaps 33;

QY 83 LKKE-NEP-----ELANISGPHKSTSTSTKRRARSKKATSDSVSKIDES----128
DB 1111 LKQVNTQNRITTELELEN-----BRQSRKAERAKSLQRELEGLDRLEQGGAT 1164

QY 129 ----VASYDSSTHLRRSSRSKKPVNYSSES-----SESEQISKATKKV 169

DB 1165 AAQVEVNNKKREAEAKLRRLDLEANNHNNHNLAAIRKKHNDVAELGDQIQEQVQKAKAKI 1224
QY 170 K-----OKEEEYVEEVDKES---LKNESSEDFEPVVPVPROLETPISKRRR-----212
DB 1225 EKDKMAQERODLVAQIDGETAERMMNNKLAQYEMQIAE-LQTKCDQONKOLQEFFTL 1283
QY 213 -SRSSAKNL-----EKSTMN:DDHAPREMPD-----CLDKPIPWGRGLGYACLN- 256
DB 1284 KTRLSNENTDLGKQIEEAESQVNMATRLKAQLTSQLEEARRSILDEARDNNLAQMKNY 1343
QY 257 -----TILSMKRVKFCSTCRITTTIQRDGL-BSVKQLQTQNVLDLIKLVENHNHFGIHF 310
DB 1344 QHEIOVLESMEFEI-----EGISELMQLSRANA-----ETQOQTFESEG 1386
QY 311 MRYSSDLTFPASHAKYGY-----TLEFAQSHLEVGKLN-----ANKVN 348
DB 1387 LLKGDLEESKKQMIKMNELQETLDAANSKISSLEKTSRLVSDLDQADQMDVERANSTY 1446
QY 349 HRLTMHPGQYQTASPRE-----VVVDSATRD-----AYHDEIL-----383
DB 1447 SQLEKQKGFQKVIDEMKKTKDDIATEVDNAQREARNVSTELFKLSEQDEVLETIEGLR 1506
QY 384 -----SRMKLNQOLN-----KDAV-----397
DB 1507 RENKELAQRKDLTDQGGGSGVFEWQKILIRLEVEKDELQHALDQPDQDAEAEESKVL 1566
QY 398 -----LIIHLGGTTEGKKTLDREKRYORLSDSVKARLVLENDVYSWSQDILLPLC 449
DB 1567 RAOVEVSQIRAEKRIQKEEDEFENTRNHQRJESMOAS--LENE--TRSKADJMLRLK 1622
QY 450 QEL-----NIPLVDHNNHIVPTLREG---SLDMLPLIPTITRETWRKGTOKOHY--499
DB 1623 KKLSDINELEIALD--HANQANAQAKNVKTYQDQMRLOQOVETEQRNRSRQYLN 1680
QY 500 -----SESADPTAISCMKRRASDRVDFPPCDPTMDLMEAKEQAVFELC---547
DB 1681 MEKATILLOQEEMSVANGQARARKQADYD-----ANEHTOCNELSAQAESLCSGR 1734
QY 548 RRYELONPPCPLEITMGPEYDQTRDGYPPGAERK-----RLTARKRSRKEEVED 597
DB 1735 RKLDTE-----LLAQADLDLETNEY--KASBEERKAASSDAARLAELQURKEQENSQND 1787

RESULT 15

S38173
myosin-like protein MLP1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YKR093w; protein YKR415
C:Species: Saccharomyces cerevisiae
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Jul-2000
C:Accession: S38173; S40647; S31207
R:Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantale submitted to the Protein Sequence Database, March 1994
A:Reference number: S38158
A:Accession: S38173
A:Molecule type: DNA
A:Residues: 1-1875 <BAL>
A:Cross-references: EMBL:Z28320; NID:g486586; PID:g486587; MIPS:YKR093w
A:Experimental source: strain S288C
R:Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; Yeast 9, 1349-1354, 1993
A:Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chr
A:Reference number: S40644; MUID:94205265
A:Accession: S40647
A:Molecule type: DNA
A:Residues: 1-1875 <BOU>
A:Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A:Experimental source: strain S288C
R:Koelling, R.; Nguyen, T.; Chen, E.Y.; Bolstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A:Title: A new yeast gene with a myosin-like heptad repeat structure.
A:Reference number: S31207; MUID:93247549
A:Accession: S31207
A:Molecule type: DNA

A;Residues: 1-300,'A',302-1875 <KOE>
A;Cross-references: EMBL:L01992; NID:g171958; PIDN:AAA34783.1; PID:g171959
C;Genetics:
A;Gene: SGD:MLP1
A;Cross-references: SGD:S0001803; MIPS:YKR095w
A;Map position: 11R

```

Query Match      4.6%; Score 144; DB 2: Length 1875;
Best local Similarity 21.4%; Pred. No. 4.6;
Matches 92; Conservative 71; Mismatches 164; Indels 102; Gaps 17;

QY  88 ETELANTISGGPHKKSTSTSTRKRARSSKKATDSYSDKIDESVASYDSTHLRRSSRSKKP 147
   : : : : : ||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1454 QNDLSNIVESMKKSFEDKTKFI ---KEKTQVNEKILAEQRLNQPSNINMEEIKKK- 1508
   : : : : : | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 148 VNYSSSESE-----SEEQISKATKKVKQKEEYVEEVDKSLKN 188
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 1509 --WESEHEQVQSQIREAEPAKKKRIPLPTEKINKTIERKKEELEKEFEKVEPER-IKS 1565
   : : : : : | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 189 ESSSDEFPVPEUPTPISKRRRSRSASAKLEKSTMNLLD-----HAPNEMFDC 239
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 1566 MEQSGEIDVILRKQLEAKVQEQ-----KELENEYKKIQEELKVDPHSHISDDERDK 1619
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
QY 240 LDKPIPWGRIGYACLANTIILRSMKERV---CSRTCRITTIQBDGLSVKQL----- 288
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 1620 LRAEIESRLREF---NNELQAIKKSGFDEGQOAMKMTLLERKLAKMESQLSQTSKQA 1676
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
QY 289 -----GTQN-VLDLILKVENHNHFGIHMRYSSDLFPFASHAKYG---YPLEPAQS 335
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 1677 ESPPKSVNNVQNPLGLPRLKRIEENSFPNPLSGEKLKLKNSKSSGGNGPFTSPSPNK 1736
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
QY 336 HLEEVG----KLANKYNRHJTMHPGQVQTQIASPVEVVDYSAIRDLAYHDEILS----- 384
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 1737 HLQNDNDKRESLANKTDPPTHLEPS--FNIPASRGLLSSSSTLSTDTNDELTNNPAQK 1794
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
QY 385 -RMKLNQLANKDAVLIILHGGTFEGK-----ETLDRFKKNYORLSDSVKARVLVN 435
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 1795 DSSNRNVQSDETT-----EKKKEGEFVKRGEAIEQOTSKNKRPIDEV-GELKND 1843
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
QY 436 DDVSWSVQD 444
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |
Db 1844 DDTTENINE 1852
   : : : : : | | | | | : : : : : | | | | | : : : : : | | | | |

```

Search completed: January 15, 2002, 13:52:22
Job time: 391 sec

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:50:21 ; Search time 134.53 seconds
(without alignments)
900.271 Million cell updates/sec

Title: US-09-724-296-2

Perfect score: 4349

Sequence: 1 MTKPLTGLGKVLQVPT.....KRLTARKRSRKEVEDEK 828

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3112	71.6	599	3 Q10988	Q10988 schizosacch
2	1200.5	27.6	262	2 Q09195	Q09195 escherichia
3	1200.5	27.6	268	2 Q09194	Q09194 escherichia
4	1181	27.2	245	2 Q09193	Q09193 escherichia
5	1181	27.2	243	2 Q09192	Q09192 escherichia
6	1181	27.2	244	2 Q09191	Q09191 escherichia
7	1043	24.0	218	5 Q94745	Q94745 schistosoma
8	1011.5	23.3	656	3 Q01408	Q01408 neurospora
9	712	16.4	218	5 Q25595	Q25595 clonorchis
10	698	16.0	218	5 Q9X199	Q9X199 fasciola gi
11	510.5	11.7	218	6 Q9N0V4	Q9N0V4 bos taurus
12	496.5	11.4	223	5 Q97117	Q97117 boophilus m
13	479	11.0	218	11 Q35660	Q35660 mus musculus
14	476.5	11.0	218	6 Q9TSM5	Q9TSM5 macaca fasc
15	476	10.9	221	6 Q9BEB0	Q9BEB0 macaca fusc
16	473.5	10.9	218	6 Q9TSM4	Q9TSM4 macaca fasc
17	468.5	10.8	219	5 Q27653	Q27653 echinococcu
18	467.5	10.7	219	5 Q16058	Q16058 echinococcu
19	460.5	10.6	218	11 Q9WU21	Q9WU21 rattus norv

20	457.5	10.5	218	11 Q9DCB8	Q9DCB8 mus musculu
21	455.5	10.5	218	11 Q9DD25	Q9DD25 mus musculu
22	446.5	10.3	225	4 Q60550	Q60550 homo sapien
23	440.5	10.1	225	6 Q9BEA9	Q9BEA9 macaca fusc
24	435	10.0	232	11 Q9D5J8	Q9D5J8 mus musculu
25	434.5	10.0	225	11 Q9Z1B2	Q9Z1B2 rattus norv
26	430.5	9.9	195	4 Q05465	Q05465 homo sapien
27	407.5	9.4	188	6 Q9MZB4	Q9MZB4 capra hircu
28	378	8.7	219	5 Q9U582	Q9U582 psoroptes o
29	359	8.3	317	2 Q9L4F0	Q9L4F0 bacillus ce
30	359	8.3	325	2 Q9L4E9	Q9L4E9 bacillus ce
31	306.5	7.0	322	2 Q9K9P8	Q9K9P8 bacillus ha
32	285.5	6.6	125	6 Q29583	Q29583 sus scrofa
33	253.5	5.8	208	5 Q02636	Q02636 brugia mala
34	253	5.8	208	13 Q9DDU5	Q9DDU5 brachydanio
35	246	5.7	210	13 P81942	P81942 bufo bufo (
36	244.5	5.6	208	5 Q27711	Q27711 onchocerca
37	240	5.5	210	4 Q15690	Q15690 homo sapien
38	240	5.5	210	4 Q00460	Q00460 homo sapien
39	236	5.4	180	5 Q9NHB2	Q9NHB2 mytilus edu
40	233	5.4	209	6 Q9TTY8	Q9TTY8 capra hircu
41	229.5	5.3	305	2 Q9S0M9	Q9S0M9 deinococcus
42	229.5	5.3	326	2 Q9RTE6	Q9RTE6 deinococcus
43	213	4.9	216	13 Q9W647	Q9W647 oncorhynch
44	212.5	4.9	82	4 Q9UE37	Q9UE37 homo sapien
45	193.5	4.4	221	11 Q9DCU1	Q9DCU1 mus musculu

ALIGNMENTS

RESULT 1
Q10988
ID Q10988 PRELIMINARY: PRT: 599 AA.
AC Q10988; P87339;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE UV-ENDONUCLEASE.
GN UVDE OR UVEL+.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sp972;
RX MEDLINE=96188860; PubMed=8614629;
RA Takao M., Yonemasa R., Yamamoto K., Yasui A.;
RT "Characterization of a UV endonuclease gene from the fission yeast
RL Nucleic Acids Res. 24:1267-1271(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97175806; PubMed=9023111;
RA Davey S., Nass M.L., Ferrer J.V., Sidik K., Eisenberger A.,
RA Mitchell D.L., Freyer G.A.;
RT "The fission yeast UVDR DNA repair pathway is inducible."
RL Nucleic Acids Res. 25:1002-1008(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRATN-972H-;
RA Lyne M., Rajandream M.A., Barrell B.G., Lucas M., Gaillardin C.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: D78571; BAA11415.1;
DR EMBL: U78487; AAC49664.1;
DR EMBL: AL023859; CAA19577.1;
KW Endonuclease.
SQ SEQUENCE 599 AA; 68816 MW; C09ED7219D5895BE CRC64;

Query Match

71.6%; Score 3112; DB 3; Length 599;

```

Best Local Similarity 100.0%; Pred. No. 6e-190;
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 MRLLLKNIQISKRIVFTILKQAFKGNHPCVSVCTIYSRFHCLPDLTKSLLPMSST 289
Db 1 MRLLLKNIQISKRIVFTILKQAFKGNHPCVSVCTIYSRFHCLPDLTKSLLPMSST 60
QY 290 TSLMPQVNIAGNSFSAETPVDLKKNETELANISGPHKKSTSTSRKARSKKKATDS 349
Db 61 TSLMPQVNIAGNSFSAETPVDLKKNETELANISGPHKKSTSTSRKARSKKKATDS 120
QY 350 VSDKIDESVASYDSSTHLRSSRSKPPVNVSSSESEEQISKATKKVQKEEYVEE 409
Db 121 VSDKIDESVASYDSSTHLRSSRSKPPVNVSSSESESEEQISKATKKVQKEEYVEE 180
QY 410 VDEKSLKNSSSEDEFEPVPEQLETPISKRRRSRSKAKLEKSTWNLDDHAPREMFDC 469
Db 181 VDEKSLKNSSSEDEFEPVPEQLETPISKRRRSRSKAKLEKSTWNLDDHAPREMFDC 240
QY 470 DKPIWGRGLGYACLNITILSRMKERVFCSTCRITTIQRDGLSVKQLGTQNVLDLIK 529
Db 241 DKPIWGRGLGYACLNITILSRMKERVFCSTCRITTIQRDGLSVKQLGTQNVLDLIK 300
QY 530 EWNHFGIHPMRYSDDLFPFASHAKYGYTFLEPAOSHLEVGKLANYNHRLTMHPGQY 589
Db 301 EWNHFGIHPMRYSDDLFPFASHAKYGYTFLEPAOSHLEVGKLANYNHRLTMHPGQY 360
QY 590 IASPREVVDSATRLAYHDEILSRMKLNEQNLKQDAVLIHLLGGTFEGKKETLDRFR 649
Db 361 IASPREVVDSATRLAYHDEILSRMKLNEQNLKQDAVLIHLLGGTFEGKKETLDRFR 420
QY 650 QRLSDSVKARLVLENDVSWSDLLPLCOELNIPVLVDWHHNIIVPGTLRGSLDLML 709
Db 421 QRLSDSVKARLVLENDVSWSDLLPLCOELNIPVLVDWHHNIIVPGTLRGSLDLML 480
QY 710 IPTRETWTIRKGTQKHYSSESADPTAISGMKRRASDRVDFPPCDPTMDLMIKAKE 769
Db 481 IPTRETWTIRKGTQKHYSSESADPTAISGMKRRASDRVDFPPCDPTMDLMIKAKE 540
QY 770 QAVFELCRRYELONPCPLEINGPEYDQTRDGYPPGAERKLTARKRRSRKEEVEDEK 828
Db 541 QAVFELCRRYELONPCPLEINGPEYDQTRDGYPPGAERKLTARKRRSRKEEVEDEK 599

RESULT 2
O09195
ID O09195 PRELIMINARY; PRT: 262 AA.
AC O09195;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE GST-6HIS.
GN GST.
OS Escherichia coli.
OC Plasmid pGEX-2T-6His-PL2.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Cahill M.A.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U84572; A841883.1; -.
DR HSP; P08515; IGNE.
DR InterPro: IPR000521; GST.
DR Pfam: PF00043; GST; 1.
SQ SEQUENCE 262 AA; 30392 MW; BA4C4593CB20226E CRC64;

Query Match 27.6%; Score 1200.5; DB 2: Length 268;
Best Local Similarity 97.8%; Pred. No. 7.4e-69;
Matches 224; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 5 PILGYWKIKGLVQPTRLLEYLEEKYEHLIERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
Db 3 PILGYWKIKGLVQPTRLLEYLEEKYEHLIERDEGDKWRNKKFELGLEFPNLPYYIDGD 62
QY 65 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYASKDFETLKVDF 124
Db 63 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYASKDFETLKVDF 122
QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHVTHPDVFLYDALDVLYMDPCLDAFPKLVCFKKR 184
Db 123 LSKLPEMLKMFEDRLCHKTYLNGDHVTHPDVFLYDALDVLYMDPCLDAFPKLVCFKKR 182
QY 185 TEAIPQIDKYLKSKYIAWPLQGWOATFGGDDHPPKSD-LVPRGSPVEM 233
Db 183 TEAIPQIDKYLKSKYIAWPLQGWOATFGGDDHPPKSD-LVPRGSPVEM 230

RESULT 3
O09194
ID O09194 PRELIMINARY; PRT: 268 AA.
AC O09194;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE GST-6HIS.
GN GST.
OS Escherichia coli.
OC Plasmid pGEX-2T-6H.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Cahill M.A.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U84571; A841882.1; -.
DR HSP; P08515; IGNE.
DR InterPro: IPR000521; GST.
DR Pfam: PF00043; GST; 1.
SQ SEQUENCE 268 AA; 31000 MW; 85820375FE60E625 CRC64;

Query Match 27.6%; Score 1200.5; DB 2: Length 268;
Best Local Similarity 97.8%; Pred. No. 7.4e-69;
Matches 224; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 5 PILGYWKIKGLVQPTRLLEYLEEKYEHLIERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
Db 3 PILGYWKIKGLVQPTRLLEYLEEKYEHLIERDEGDKWRNKKFELGLEFPNLPYYIDGD 62
QY 65 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYASKDFETLKVDF 124
Db 63 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYASKDFETLKVDF 122
QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHVTHPDVFLYDALDVLYMDPCLDAFPKLVCFKKR 184
Db 123 LSKLPEMLKMFEDRLCHKTYLNGDHVTHPDVFLYDALDVLYMDPCLDAFPKLVCFKKR 182
QY 185 TEAIPQIDKYLKSKYIAWPLQGWOATFGGDDHPPKSD-LVPRGSPVEM 233
Db 183 TEAIPQIDKYLKSKYIAWPLQGWOATFGGDDHPPKSD-LVPRGSPVEM 230

RESULT 4
O09193
ID O09193 PRELIMINARY; PRT: 245 AA.
AC O09193;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)

```

01-JUN-2001 (TREMELREL. 17, Last annotation update)
DE GLUTATHIONE S-TRANSFERASE.
DN
OS Escherichia coli.
OG Plasmid pGEX-6p-3.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88329742; PubMed=3047011;
RA Smith D.B., Johnson K.S.;
RT "Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase.";
RL Gene 67:31-40(1988).
RN [2]
RP SEQUENCE OF 221-227 FROM N.A.
RX MEDLINE=90264387; PubMed=2160953;
RA Cordingley M.G., Callahan P.L., Sardana V.V., Garsky V.M.,
Colanno R.J.;
RT "Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro.";
RL J. Biol. Chem. 265:9062-9065(1990).
RN [3]
RP SEQUENCE FROM N.A.
RA Dunst R.W., Bell P.A., English J.D.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U78874; AAB37352.1; -;
DR HSSP: P08515; IGNE.
DR InterPro: IPR000521; GST.
DR Pfam: PF00043; GST; 1.
KW Transferase; Plasmid.
SQ SEQUENCE 245 AA; 28327 MW; D51470E8A48C6CC2 CRC64;

Query Match 27.2%; Score 1182; DB 2; Length 245;
Best Local Similarity 92.2%; Pred. No. 9.8e-68;
Matches 226; Conservative 1; Mismatches 10; Indels 8; Gaps 2;

QY 5 PILGYWKIKGLVQPTRLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
Db 3 PILGYWKIKGLVQPTRLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLPYYIDGD 62

QY 65 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRAYSKDFTLKVD 124
Db 63 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRAYSKDFTLKVD 122

QY 125 LSKLPEMLKMFEDRLCHTKTYLNGDHVTHPDFMLYDALDVVLYMDPCLDAFPKLVCFKKR 184
Db 123 LSKLPEMLKMFEDRLCHTKTYLNGDHVTHPDFMLYDALDVVLYMDPCLDAFPKLVCFKKR 182

QY 185 IEAIPQIDKYLKSSKYIAWPLOGWOATFGGDHPPKSDHLV----PRGSMRLRLKRNIOI 240
Db 183 IEAIPQIDKYLKSSKYIAWPLOGWOATFGGDHPPKSDHLV----NSRVDS 238

QY 241 SKRIV 245
Db 239 SGRIV 243

RESULT 5
009192 ID 009192 PRELIMINARY; PRT: 243 AA.
AC 009192;
DT 01-JUL-1997 (TREMELREL. 04, Created)
DT 01-JUL-1997 (TREMELREL. 04, Last sequence update)
DT 01-JUN-2001 (TREMELREL. 17, Last annotation update)
DE GLUTATHIONE S-TRANSFERASE.
GN GST.
OS Escherichia coli.
OG Plasmid pGEX-6p-2.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.

OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88329742; PubMed=3047011;
RA Smith D.B., Johnson K.S.;
RT "Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase.";
RL Gene 67:31-40(1988).
RN [2]
RP SEQUENCE OF 221-227 FROM N.A.
RX MEDLINE=90264387; PubMed=2160953;
RA Cordingley M.G., Callahan P.L., Sardana V.V., Garsky V.M.,
Colanno R.J.;
RT "Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro.";
RL J. Biol. Chem. 265:9062-9065(1990).
RN [3]
RP SEQUENCE FROM N.A.
RA Dunst R.W., Bell P.A., English J.D.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U78873; AAB37349.1; -;
DR HSSP: P08515; IGNE.
DR InterPro: IPR000521; GST.
DR Pfam: PF00043; GST; 1.
KW Transferase; Plasmid.
SQ SEQUENCE 243 AA; 27908 MW; 0003C434CD76C3A6 CRC64;

Query Match 27.2%; Score 1181; DB 2; Length 243;
Best Local Similarity 96.9%; Pred. No. 1.1e-67;
Matches 222; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

QY 5 PILGYWKIKGLVQPTRLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
Db 3 PILGYWKIKGLVQPTRLLEYLEEKYEHLRYERDEGDKWRNKKFELGLEFPNLPYYIDGD 62

QY 65 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRAYSKDFTLKVD 124
Db 63 VKLTQSMALIRYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRAYSKDFTLKVD 122

QY 125 LSKLPEMLKMFEDRLCHTKTYLNGDHVTHPDFMLYDALDVVLYMDPCLDAFPKLVCFKKR 184
Db 123 LSKLPEMLKMFEDRLCHTKTYLNGDHVTHPDFMLYDALDVVLYMDPCLDAFPKLVCFKKR 182

QY 185 IEAIPQIDKYLKSSKYIAWPLOGWOATFGGDHPPKSDHLV----PRGS 229
Db 183 IEAIPQIDKYLKSSKYIAWPLOGWOATFGGDHPPKSDHLV----PRGS 231

RESULT 6
009191 ID 009191 PRELIMINARY; PRT: 244 AA.
AC 009191;
DT 01-JUL-1997 (TREMELREL. 04, Created)
DT 01-JUL-1997 (TREMELREL. 04, Last sequence update)
DT 01-JUN-2001 (TREMELREL. 17, Last annotation update)
DE GLUTATHIONE S-TRANSFERASE.
GN GST.
OS Escherichia coli.
OG Plasmid pGEX-6p-1.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88329742; PubMed=3047011;
RA Smith D.B., Johnson K.S.;
RT "Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase.";
RL Gene 67:31-40(1988).
RN [2]
RP SEQUENCE OF 221-227 FROM N.A.
RX MEDLINE=90264387; PubMed=2160953;

Qy	656	VKARVLVENDDVSNVSODLLPCLQEENLPIVLVDHWHNTV--PGTLREGSLDIM--PLIP	711
Dd	414	CNRLRVLENDVGWTHDLLPVCEELINPMVLDYHHNNICFDPFAHLREGTLDISDPKLOE	473
Qy	712	TIRETWTRKGITOKOYSESADPTALSGMKRRASHSDRVDFPCDPTMDLMTEAKEQA	771
Dd	474	RIANTWRKGIKOMHYSEPCD--GAVTPDRRKRHRVMTLPPCPDMDLMTEAKDEQA	532
Qy	772	VFELCRRYEL-----QNPPCP-----LETMGE---794	
Dd	533	VFELMRTFKLPGFEKINDMVPYDRDDENRAPVPKPKKKGGKGRKTDTDEAAPEEVD	592
Qy	795	--YDQPDG-----YTPPGAERLTARKRSRK-----EEVEED	826
Dd	593	TAADDVKDADEGPKEVPEERAMGGPYNRVYPVLPGLCEEWLKPKREVRKGVPEEVED	651

RESULT	9	
Q25595		
ID	PRELIMINARY;	PRT; 218 AA.
AC	Q25595;	
DT	01-NOV-1996 (TrEMBLrel. 01, Created)	
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)	
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)	
DE	PUTATIVE GLUTATHIONE TRANSFERASE.	
OS	Clonorchis sinensis.	
OC	Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;	
OC	Trematoda; Digenea; Opisthorchiida; Opisthorchioidea; Opisthorchioidea;	
OC	Opisthorchiidae; Clonorchis.	
OX	NCBI_Taxid:79923;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Hong S.-J.; Lee D.-H.;	
RT	"Cloning and over expression of 26 kDa glutathione S-transferase from	
RT	Clonorchis sinensis."	
RL	Thesis (1996), Parasitology, Chung-Ang University College of Medicine	
RL	Seoul.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RA	Hong S.-J.;	
RL	Submitted (APR-1999) to the EMBL/GenBank/DBDJ databases.	
DR	EMBL; L47992; AAB46369.3; .	
DR	HSSP; P31670; 1FHE.	
DR	InterPro: IPR000521; GST.	
DR	Pfam: PF00043; GST; 1.	
DR	Transferase.	
KW	SEQUENCE	218 AA; 25038 MW; 7CB17C7B837A0B7C CRC64;

[illegible]

RESULT	10
Q9XYL9	

ID	Q9XYL9	PRELIMINARY;	PRT;	218	AA.
AC	Q9XYL9;				
DT	01-NOV-1999	(TrEMBLrel. 12, Created)			
DT	01-NOV-1999	(TrEMBLrel. 12, Last sequence update)			
DT	01-JUN-2001	(TrEMBLrel. 17, Last annotation update)			
DE	GLUTATHIONE S-TRANSFERASE.				
GN	GST-1.				
OS	Fasciola gigantica.				
OC	Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;				
OC	Trematoda; Digenea; Echinostomida; Echinostomata; Fascioloidae;				
OC	Fasciolidae; Fasciola.				
OX	NCBI_TaxID=46835;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=THAILAND;				
RA	Grams S.V., Grams R., Sobhon P., Viyanant V., Upatham E.S.;				
RT	"Best molecular cloning of expressed antigens from Fasciola gigantica.";				
RL	submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF112567; AAD23997.1;				
DR	HSSP; P31670; 1PHE.				
DR	InterPro; IPR00521; GST.				
DR	Pfam; PF00043; GST; 1.				
KW	Transferase.				
SC	SEQUENCE	218	AA;	25337	MW; E9423D75C3F2EEAF CRC64;
Query Match 16.0%; Score 698; DB 5; Length 218;					
Best Local Similarity 58.7%; Pred. No. 5.1e-37;					
Matches 125; Conservative 35; Mismatches 53; Indels 0; Gaps					
Qy	7	LGYNKIKGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGLFEPNLPYYIDGDVK	66		
Db	5	LGYNKIRGLQPPVRLLEYLEEYEEHLYGRDDREKWLGDKNPGLDLPNLPYYIDDKCK	64		
Qy	67	LTQSMALIRIADKHNMLGCGPKERARISMLEGAVLDIRYGVSRVSIAYSKDFETLKVDFLS	126		
Db	65	LTQSVAMIRIADKHGMLGSTPEERARVSIIEGAAMDRLRMCFVRVCYNPNFEEVKGDKL	124		
Qy	127	KLPKLMKFEDRLCHKTYLNGDHWTHPDFMLYDALDVLVYMDPMLDAPFKLVCFKKRIE	186		
Db	125	ELPKTLKMSDFGLGROYLTGSSVSHVDFVYALDCIRYLAPQCLNDFPKLKEFKSRIE	184		
Qy	187	AIPOIDRYLKSSYIAMPLOGWQATFGGGHPP	219		
Db	185	DLPIKAIYMESEKFIKWLPLNSWTASFGGGDAAP	217		
RESULT	11				
Q9NOV4					
ID	Q9NOV4	PRELIMINARY;	PRT;	218	AA.
AC	Q9NOV4;				
DT	01-OCT-2000	(TrEMBLrel. 15, Created)			
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)			
DT	01-JUN-2001	(TrEMBLrel. 17, Last annotation update)			
DE	CLASS MU GLUTATHIONE S-TRANSFERASE.				
GN	GSTM.				
OS	Bos taurus (Bovine).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;				
OC	Bovidae; Bovinae; Bos.				
OX	NCBI_TaxID=9913;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=EYE LENS;				
RA	Jimenez-Asensio J.V., Gariand D.;				
RT	"A lens glutathione S-transferase, class mu, with thiol-specific				
RL	antioxidant activity.";				
DR	submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF249588; AAF64308.1;				
DR	InterPro; IPR000521; GST.				
DR	InterPro; IPR003081; GST_mu.				
DR	Pfam; PF00043; GST; 1.				
DR	PRINTS; PR01267; GSTRNSFRASEM.				

```

KW Transferase.
SQ SEQUENCE 218 AA; 25635 MW; 3D02EA0F43C07B0A CRC64;

Query Match 11.7%; Score 510.5; DB 6; Length 218;
Best Local Similarity 47.8%; Pred. No. 4.2e-25;
Matches 98; Conservative 35; Mismatches 67; Indels 5; Gaps 1;

QY 6 ILGYWKIKGLVOPTRLLLEYLEEKEEHLRYERDEG-----DKWRNKKFELGLEFPNLY 60
DB 4 ILGYWDIRGLHAIRLLLEVTDTNTEERQYSVGDADYDRSQWLNEKFKLGLDFNPLYL 63

QY 61 IDGDVKLTOSMATIRYADKHNMLGCPKRAEISMLEGAVLDIRYGVSRAYSXDFETL 120
DB 64 IDGTHKLTOSNATLRYAIKHNLCGTEEMIRVDILENOVMDVRLAMARICYSDFPEKL 123

QY 121 KVDPLSKPLMKMFEDRLCHTKTYLNGDRVTHPHDFMLYDALDVLVYMDPMLCLDAPFKLVC 180
DB 124 KPGFLAKEIPEKIFSEFLGKRPWFAGDKLTYVDVFLVYDVLDMHRTFEPKCLDAFNLKD 183

QY 181 FKRIEAIPOIDKYLKSSKYYIAWPL 205
DB 184 FISRFELKKISAYMKSSRFLPGPL 208

RESULT 12
ID 097117 PRELIMINARY: PRT; 223 AA.
AC 097117;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE GLUTATHIONE S-TRANSFERASE.
OS Boophilus microplus (Cattle tick).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Parasitiformes; Ixodida; Ixodidae; Ixodidae; Boophilus.
OX NCBI_TaxID=6941;
RN [1]
RP SEQUENCE FROM N.A.
RA He H., Chen A.C.;
RT "Characterization and molecular cloning of a glutathione S-transferase
RT from larvae of the cattle tick Boophilus microplus (Acari;
RT Ixodidae).";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077609; AAD15991.1; -.
DR HSP; P20136; IGSL.
DR InterPro; IPR000521; GST.
DR InterPro; IPR003081; GST_mu.
DR Pfam; PF000043; GST; 1.
DR PRINTS; PR01267; GSTRNSFRASEM.
KW Transferase.
SQ SEQUENCE 223 AA; 25604 MW; F276ACE2D8259174 CRC64;

Query Match 11.4%; Score 496.5; DB 5; Length 223;
Best Local Similarity 47.2%; Pred. No. 3.4e-24;
Matches 101; Conservative 28; Mismatches 80; Indels 5; Gaps 1;

QY 5 PILGYWKIKGLVOPTRLLLEYLEEKEEHLRYERD-----EGDKWRNKKFELGLEFPNLY 59
DB 3 PVLGYWDIRGLAQPIRELLAHVDAKVDKRYTCGPPDFDRSSWLNEKTKLGLFEPNLY 62

QY 60 YIDGDVKLTOSMATIRYADKHNMLGCPKRAEISMLEGAVLDIRYGVSRAYSXDFET 119
DB 63 YIDGDVKLTOSMATIRYLARKHLGLEGTEAEKORVDVSEQQQAFDFRNWVRLCYNDFEK 122

QY 120 LKVDPLSKPLMKMFEDRLCHTKTYLNGDRVTHPHDFMLYDALDVLVYMDPMLCLDAPFKLV 179
DB 123 LKGDYLNLPASLAFSDYLGTHKFFAGDNLTYVDFTAYEMLAQHLIFADCLKDFANLK 182

QY 180 CFKKRIEAIPOIDKYLKSSKYYIAWPLQWQATFG 213
DB 183 AFVDRIALPHVAAYLKSDCKIKWPLNGDMASFG 216

```

```

RESULT 13
ID 035660 PRELIMINARY: PRT; 218 AA.
AC 035660; 035661;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE GLUTATHIONE S-TRANSFERASE 6 (EC 2.5.1.18) (GLUTATHIONE-S-TRANSFERASE
DE CLASS M5).
GN GSTM6 OR GSTM5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN-FVB, 129/SV, AND C57BL/6; TISSUE=LIVER;
RX MEDLINE=98149729; PubMed=9480867;
RA De Bruin W.C.C., te Morsche R.H.M., Wagenmans M.J.M., Alferink J.C.,
RA Townsend A.J., Wieringa B., Peters W.H.M.;
RT "Identification of a novel murine glutathione S-transferase class mu
RT gene.";
RL Biochem. J. 330:623-626(1998).
CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER OF
CC EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-G.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, STOMACH AND SMALL
CC INTESTINE. NOT EXPRESSED IN SPLEEN, KIDNEY, COLON, HEART, MUSCLE,
CC BRAIN OR LUNG.
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
DR EMBL; AJ000413; CAAG4061.1; -.
DR EMBL; AJ000412; CAAG4060.1; -.
DR HSP; P09488; IGTU.
DR MGD; MGI:1309467; Gstm6.
DR InterPro; IPR000521; GST.
DR InterPro; IPR003081; GST_mu.
DR Pfam; PF000043; GST; 1.
DR PRINTS; PR01267; GSTRNSFRASEM.
KW Transferase; Multigene family.
SQ SEQUENCE 218 AA; 25628 MW; 324B20CC6C35594B CRC64;

Query Match 11.0%; Score 479; DB 11; Length 218;
Best Local Similarity 44.0%; Pred. No. 4.3e-23;
Matches 95; Conservative 37; Mismatches 78; Indels 6; Gaps 2;

QY 4 LPI-LGYWKIKGLVOPTRLLLEYLEEKEEHLRYERDEG-----DKWRNKKFELGLEFPNL 57
DB 1 MPVTLGYWDIRGLGHAIIRLLLEVTDTNTEERQYSVGDADYDRSQWLNEKFKLGLDFPNL 60

QY 58 PYIDGDVKLTOSMATIRYADKHNMLGCPKRAEISMLEGAVLDIRYGVSRAYSXDFE 117
DB 61 PVLIDGSHKVTQSNAILRYLGRKHNLCGTEEERIRVDILENRYMDTRIOMGLCYXADF 120

QY 118 ETLKVDPLSKPLMKMFEDRLCHTKTYLNGDRVTHPHDFMLYDALDVLVYMDPMLCLDAPPK 177
DB 121 EKKRKEFLKGLPDQLKLYSEFLGQWPFAGDKITFADFLVYDVLDOHRMFEPTCLDAPFN 180

QY 178 LVCFKKRIEAIPOIDKYLKSSKYYIAWPLQWQATFG 213
DB 181 LKDFMARFELGKIKISAYMKTSRFLSPYLKQATWG 216

RESULT 14
ID 097SM5 PRELIMINARY: PRT; 218 AA.
AC 097SM5;
DT 01-MAY-2000 (Tremblrel. 13, Created)

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:06:31 ; Search time 134.53 Seconds
(without alignments)
319.662 Million cell updates/sec

Title: US-09-724-296-39

Perfect score: 1527
Sequence: 1 QLGLVCLTVGPEVRFTVTLL.....VEAKGKEAIAALRLMAPFK 294

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL_17:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_Organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_prodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1520	99.5	305	2 Q9S0M9	Q9S0M9 deinococcus
2	1520	99.5	326	2 Q9RTE6	Q9RTE6 deinococcus
3	297.5	19.5	317	2 Q9L4F0	Q9L4F0 bacillus ce
4	297.5	19.5	325	2 Q9L4E9	Q9L4E9 bacillus ce
5	262.5	17.2	656	3 Q01408	Q01408 neurospora
6	250	16.4	322	2 Q9K9P8	Q9K9P8 bacillus ha
7	230.5	15.1	599	3 Q10988	Q10988 schizosacch
8	111	7.3	315	2 Q9AMN9	Q9AMN9 pseudomonas
9	102.5	6.7	437	2 Q9PFJ8	Q9PFJ8 xylella fas
10	98.5	6.5	313	1 Q9VAQ7	Q9VAQ7 aeropyrum p
11	98.5	6.5	1486	2 Q9CG24	Q9CG24 lactococcus
12	96.5	6.3	2216	10 Q9LIV03	Q9LIV03 arabidopsis
13	96	6.3	526	2 Q9X5N3	Q9X5N3 myxococcus
14	96	6.3	535	2 Q9GCB8	Q9GCB8 lactococcus
15	96	6.3	791	10 Q90962	Q90962 arabidopsis
16	95	6.2	368	2 Q9XAC4	Q9XAC4 streptomyce
17	95	6.2	515	2 Q9RYP5	Q9RYP5 deinococcus
18	95	6.2	558	11 Q9JL61	Q9JL61 mus musculus
19	94.5	6.2	388	2 Q9KID6	Q9KID6 streptomyce

20	94.5	6.2	1203	2 Q9ACX5	Q9acx5 streptomyce
21	94.5	6.2	1809	2 Q85740	Q85740 pseudomonas
22	94.5	6.2	1809	2 Q9HWG4	Q9hwg4 pseudomonas
23	94.5	6.2	2731	2 Q923T9	Q923t9 pseudomonas
24	94	6.2	898	2 Q9RI40	Q9ri40 streptomyce
25	93.5	6.1	1809	2 Q9RFM7	Q9rfm7 pseudomonas
26	93	6.1	603	2 Q9RTX0	Q9rtx0 deinococcus
27	92.5	6.1	915	11 Q9ERV7	Q9erv7 mus musculi
28	92	6.0	332	2 Q9ZGA5	Q9zga5 streptomyce
29	92	6.0	777	2 Q9AAW1	Q9aaw1 caulobacter
30	92	6.0	951	2 Q9HTC4	Q9htc4 pseudomonas
31	92	6.0	1017	4 Q9V2M0	Q9v2m0 homo sapien
32	91.5	6.0	6315	2 Q9ADL6	Q9adl6 polyangium
33	91	6.0	344	2 Q9I5Q9	Q9i5q9 pseudomonas
34	91	6.0	750	2 Q9FC09	Q9fc09 streptomyce
35	90.5	5.9	586	2 P94134	P94134 alcaligenes
36	90.5	5.9	644	10 Q9ARL6	Q9arl6 hordeum vul
37	90.5	5.9	842	2 Q9HXN5	Q9hxn5 pseudomonas
38	90.5	5.9	2040	5 Q9V696	Q9v696 drosophila
39	90.5	5.9	4247	2 Q9L8H4	Q9l8h4 streptomyce
40	90.5	5.9	11096	2 Q9L4W3	Q9l4w3 streptomyce
41	90	5.9	1421	2 Q9L8C9	Q9l8c9 polyangium
42	90	5.9	3413	2 Q54593	Q54593 amycolatops
43	89.5	5.9	319	2 Q9HZ95	Q9hgz95 pseudomonas
44	89.5	5.9	390	2 Q9KCE3	Q9kce3 bacillus ha
45	89.5	5.9	397	2 Q9S2R3	Q9s2r3 streptomyce

ALIGNMENTS

RESULT 1

Q9S0M9 PRELIMINARY: PRT: 305 AA.
ID Q9S0M9
AC Q9S0M9;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE UV-ENDONUCLEASE.
GN UVSCDE.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KRI;
RA Kitayama S., Kikuchi M., Funayama T., Narumi I., Watanabe H.;
RT "Cloning of structural gene of an alternative incision enzyme for DNA
damage in Deinococcus radiodurans."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB033747; BAA85759.1; .
KW Endonuclease.
SQ SEQUENCE 305 AA; 33592 MW; B94D333243E2FEA4 CRC64;

Query Match 99.5%; Score 1520; DB 2; Length 305;
Best Local Similarity 99.7%; Pred. No. 3.2e-114;
Matches 293; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	QLGLVCLTVGPEVRFTVTLSRYRALSPAEREAKLLDLYSSNIKTLRGAADYCAAHDIRL	60
DB	10	QLGLVCLTVGPEVRFTVTSRYRALSFAEREAKLLDLYSSNIKTLRGAADYCAAHDIRL	69
QY	61	YRLSSLPFMLDLACDDTGAAVTLHAPOLLEAGHAFTDAGVRLMHPEQFIVLNSDRPE	120
DB	70	YRLSSLPFMLDLACDDTGAAVTLHAPOLLEAGHAFTDAGVRLMHPEQFIVLNSDRPE	129
QY	121	VRESSVRAMSAHARYMDGLGLARTPNLLHGGKGGCAELAAALIPDLPPVRLRLGLE	180
DB	130	VRESSVRAMSAHARYMDGLGLARTPNLLHGGKGGCAELAAALIPDLPPVRLRLGLE	189
QY	181	NDERAYSPELLPTICATGTTPLFDDAHVHVDKLPDQEDFPSVREWVLRARATWOPPEWQ	240

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 NCBI_TaxID=3702;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=20181125; PubMed=10718197;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
 RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
 RT clones.";
 RL DNA Res. 7:31-63(2000).
 CC -1- COFACTOR: FAD (BY SIMILARITY).
 CC -1- SIMILARITY: TO PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES
 CC CLASS=1.
 CC EMBL; AB020754; BAA97323.1; -;
 DR InterPro: IPR000759; Adrnx_reductase.
 DR InterPro: IPR002489; DUF14.
 DR InterPro: IPR001327; FAD_pyr_redox.
 DR InterPro: IPR002932; Glu_synthase.
 DR InterPro: IPR000205; NAD_binding.
 DR InterPro: IPR00103; Pyridine_redox_2.
 DR InterPro: IPR001100; pyr_redox.
 DR Pfam; PF01493; DUF14; 1.
 DR Pfam; PF01645; Glu_synthase; 1.
 DR Pfam; PF00070; pyr_redox; 1.
 DR PRINTS; PR00419; ADXROTASE.
 DR PRINTS; PR00368; FADPNR.
 DR PRINTS; PR00411; PNDROTASE1.
 DR PRINTS; PR00469; PNDROTASEII.
 KW FAD; Flavoprotein; Oxidoreductase; Redox-active center.
 SQ SEQUENCE 2216 AA; 242877 MW; 2AF258AB8C682371 CRC64;

Query Match 6.3%; Score 96.5; DB 10; Length 2216;
 Best Local Similarity 22.0%; Pred. No. 72;
 Matches 69; Conservative 54; Mismatches 116; Indels 75; Gaps 17;

QY 14 RFRVTLRYALSPAEREAKLDLYSSNKTLRGAADYCAAH--LRLYRLSS-----66
 Db 332 RFSNTTPSWDRAP-----NRVLG-HNGEINTLRGNVNRAREGLKCNELGSKELK 386
 QY 67 -LFPMDLADGDDTGAVALTHLAPOLLEAGHAFTDAGVRLMHPEQFIVLNSDR--PEVRE 123
 Db 387 KLLPIVDVSSDSGA--FDGVLELLVRAGRSLPEA--VMMNIPEAW---QNDKNIDPSRK 439
 QY 124 SSVRAMSAHARVMDGLGL-----ARTPWMLLLHGKGGRGGAELALIP 167
 Db 440 EFEYLSALMEPWDPGALISFTDGRYLGATLDNRGLRPGREFYTHSGRVIMASEVG--VV 497
 QY 168 DLDPVRLRLGLENDERAYSPAELLPTCEATGTPLVFDAAHHVHVHDKLPQEDPSPVREW 227
 Db 498 DVPPEDVMKGRNL-----PGMILLVDFEKHIVD-----DDLKQOY 535
 QY 228 LRATWQPEWQVHLSNIEG-PQDRRHSHLIADPPSAYADVPQ---IEVEAKKEEA 283
 Db 536 SLARPYGEWLKROKLEKDIIESVPEAE--IA--PSISGVVPLTFLETKASNDNDDS 589
 QY 284 IAAL---RLMAPFK 294
 Db 590 MESMGHGLLSPLK 603

RESULT 13
 Q9X5N3 PRELIMINARY; PRT; 526 AA.
 AC Q9X5N3;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE OLIGOPEPTIDE PERMEASE HOMOLOG OPPA.
 GN OPPA.
 OS Myxococcus xanthus.
 OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
 OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
 NCBI_TaxID=34;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=DZFL;
 RA Ueki T., Inouye S.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF127082; AAD31006.1; -;
 DR InterPro: IPR000914; SBP_bac_5.
 DR Pfam; PF00496; SBP_bac_5; 3.
 SQ SEQUENCE 525 AA; 58593 MW; AFA585E37B207DB3 CRC64;

Query Match 6.3%; Score 96; DB 2; Length 526;
 Best Local Similarity 26.6%; Pred. No. 12;
 Matches 58; Conservative 25; Mismatches 81; Indels 54; Gaps 11;

QY 18 VTLRYALSPAEREAKLDLYSSNKTLRGAADY-----52
 Db 135 VTLARPSYFLA-RVANYVLFYPAPSDLEGKSDSEVDFDRDRGRPLALGPYRVERW 193
 QY 53 -CAAHDIRLYRLSSSLFPMDLADGDDTGAVALT---HLAPQLF---AGHAFTDAGVRL 104
 Db 194 DRAGERVRLVHNPASAFP-RMAEGETVPVITLMKSEIGALYERERVDVFDVDSAAAL 252
 QY 105 LMH-PEQ-----FIVLNS-----DRPEVRESSVRAHSAHARVMDGLGLARTPNWL 148
 Db 253 RIHRPDDLRRREPLLSTYFLAFNTEKAPLDRPEVRRLSALDREA-LLAGLLPAARPSHV 311
 QY 149 LLLHGKGGSGGAELALIP-LPDPVRLRL-GLENDER 184
 Db 312 LLPPELGAATPEQAARLPYAPAEQARAEALAGVERPLR 349

RESULT 14
 Q9CG88 PRELIMINARY; PRT; 535 AA.
 AC Q9CG88;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HYPOTHETICAL PROTEIN YLJF.
 GN YLJF.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Lactococcus.
 NCBI_TaxID=1360;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=IL1403;
 RA Holstin A., Wincker P., Manger S., Jaillon O., Malarne K.,
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus
 RT lactis.";
 RL Genome Res. 0:0-0(2001).
 CC -1- COFACTOR: FAD (BY SIMILARITY).
 DR EMBL; AE006350; AAK05277.1; -;
 DR InterPro: IPR001327; FAD_pyr_redox.
 DR PRINTS; PR00368; FADPNR.
 KW Complete proteome; FAD; Flavoprotein; Hypothetical protein;
 KW Oxidoreductase.
 SQ SEQUENCE 535 AA; 58670 MW; 116102ABB9FAEC95 CRC64;

Query Match 6.3%; Score 96; DB 2; Length 535;
 Best Local Similarity 24.4%; Pred. No. 12;
 Matches 68; Conservative 36; Mismatches 87; Indels 88; Gaps 17;

QY 11 PEVRETVTLRYALSPAEREAKLD--LYSSNKTLRGAADYCAAHDIRL---YRLS 64

Db 390 --ESPTD-----LLA-ISSGYS-----HDNESHEKOKIEARRRL 422

Search completed: January 15, 2002, 14:06:33
Job time: 972 sec

Db 14 PVSXKSLVKKLRL-----QESDLDYRIYKESIDARHGEIDFYVDIKLDEARIL 68
QY 65 S-----SLFPMLDLAGDDTGAVALTHLAPQLLEAGHAFTDAGVRLLMHPEQFIVLNSDR 118
Db 69 SKKIKNVSLAPELDYVNPVIGSKKMLH-RPVI-----GFGPAG-----MFAALLAQNGYR 119
QY 119 PEVRESSVRAMSAHARVMDGLGLARTPWNLILLH-----GKGGRG-----AELAA 164
Db 120 PIVLERG-QAVDERVKSIDKF-----W-----LEGKLNPKSNVOFGEGGAGTFSDGKLT 168
QY 165 LIPDLPVRLRLGLENDERAYSPAELLIPCEATGTPLVEDAHVHVHDKLPD----- 217
Db 169 RVRDL-----RGRKVLEEFVQAGAPEDIL-----YKAHPHVCTDLRLDIVKNIRK 213
QY 218 -----QEDPSVREWLVRARATWQPPVQVHLNSG 247
Db 214 QIIELGGEVHFDAQVEEFLI-----EDELQAVKLADG 246

RESULT 15

O80962 PRELIMINARY; PRT; 791 AA.
AC O80962;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE T16B24.17 PROTEIN.
GN T16B24.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Rounsley S.D., Kaul S., Lin X., Ketchum K.A., Crosby M.L.,
RA Brandon R.C., Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D.,
RA Somerville C.R., Venter J.C.;
RT "Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004697; AAC28988.1; -
DR InterPro; IPR000209; Peptidase_S8.
DR PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN_1.
SQ SEQUENCE 791 AA; 88694 MW; C771BC05757D2863 CRC64;

Query Match 6.3%; Score 96; DB 10; Length 791;
Best Local Similarity 24.1%; Pred. No. 20;
Matches 83; Conservative 30; Mismatches 107; Indels 124; Caps 18;
QY 36 LDLYSSNITLIRGAADYCAAHDIRLYRLSSLLFPMLDLAGDDTGAVALTHLAP-----QL 90
Db 114 VDLFMADVKALETYAGY-----FYSLSKMSRPLPEVDSQAVDYENCRPHWAFRL 166
QY 91 LEAGHAFTDAGVRL-----LMHPEQ-----FIVLN---SDRPEV 121
Db 167 LEVFAFTIAAIRLRTSASVKGSLEASEQIFGMVLMHGLGPTFTKVGQSLSRPTDI 226
QY 122 --RESSVRAMSAHARVMDGLGLARTPW--NLLLLHGGKG----- 157
Db 227 IGTEISKALSELHRI-----PPPWPEAVKILEGELGGPVSEFFSOFQSVTAASFG 280
QY 158 -----RGAELAAI--POLPDVPR-----LRGL-----ENDERAYS----- 187
Db 281 QYVRCRTLDGADVAVKQRPDLRHAVLRDIVILRLGLVLRKAKRENDIRVYADELCMG 340
QY 188 -PAELLPTCEATGTPLVDFAHHVHVHDKLPDQEDPSVREWLVRARATWQPPVQVHLNS 246
Db 341 LAGELDTLEANASEFQEAHAREFSYIRPKVYQHLTKRKRLTM-----EMVVG----- 389
QY 247 GTEGQDRRHSHLIADFTFSAYADVQIEVEAKGKEEIAALRLM 290

No PA

How does this work?
how to handle?

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: January 15, 2002, 13:52:37 ; Search time 74.94 Seconds
(without alignments)
298.843 Million cell updates/sec
Title: US-09-724-296-39
Perfect score: 1527
Sequence: 1 QLGLVCLTVGPEVRFRTVTLS.....VEAKGKEEAIALRLMAPFK 294
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1520	99.5	326	2 C75350	probable UV damage
2	303	19.8	320	2 S55418	UV-endonuclease ho
3	255.5	16.7	656	2 S55262	UV-endonuclease -
4	250	16.4	322	2 B83974	hypothetical prote
5	230.5	15.1	599	2 S71134	UV-endonuclease -
6	102.5	6.7	437	2 B82778	cell cycle protein
7	98.5	6.5	313	2 F72575	hypothetical prote
8	98.5	6.5	1486	2 F86785	glutamate synthase
9	96	6.3	535	2 C82772	conserved hypotet
10	96	6.3	791	2 T02583	probable ABC trans
11	95	6.2	368	2 T36004	hypothetical prote
12	95	6.2	515	2 H75579	hypothetical prote
13	94.5	6.2	1809	2 T17403	pyochelin syntheta
14	94.5	6.2	1809	2 C83118	pyochelin syntheta
15	94	6.2	898	2 T37037	pyruvate phosphata
16	93	6.1	199	2 JU0052	puromycin N-acetyl
17	93	6.1	603	2 A75373	probable N-acetyl
18	92	6.0	332	2 T17427	probable FK506 oxl
19	92	6.0	931	2 G82965	conserved hypotet
20	91	6.0	344	2 E83562	N-acetyl-gamma-glu
21	90.5	5.9	457	2 T10360	late expression fa
22	90.5	5.9	842	2 C83177	probable phosphotr
23	90	5.9	488	2 S27652	probable aldehyde
24	90	5.9	608	1 S01040	glutamine--fructos
25	90	5.9	3413	2 T17467	rifamycin polyketi
26	89.5	5.9	319	2 H83253	conserved hypotet
27	89.5	5.9	390	2 D83853	cystathionine beta
28	89.5	5.9	397	2 T35289	probable multifunc
29	89.5	5.9	1602	2 H70984	probable polyketid

ALIGNMENTS

RESULT 1
C75350
probable UV damage endonuclease - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: C75350
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: C75350
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-326 <WHI>
A:Cross-references: GB:AE0002022; GB:AE000513; NID:g6459590; PIDN:AAF11370.1; PTD:g645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1819
A:Map position: 1
Query Match 99.5%; Score 1520; DB 2; Length 326;
Best Local Similarity 99.7%; Pred. No. 2.3e-122;
Matches 293; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	QLGLVCLTVGPEVRFRTVTLSRYRALSPEAREAKLLDLYSSNIKTILRGADYCAAHDIRL	60
Db	31	QLGLVCLTVGPEVRFRTVTLSRYRALSPEAREAKLLDLYSSNIKTILRGADYCAAHDIRL	90
QY	61	YRLSSSLFPMILDAGDDTCAAVLTHLAPOLLGAGHAFTDAGVRLLMHPQFIVLNSDRPE	120
Db	91	YRLSSSLFPMILDAGDDTCAAVLTHLAPOLLGAGHAFTDAGVRLLMHPQFIVLNSDRPE	150
QY	121	VRESSVRAMSAHARVMDGLGLARTPNWLLHGGKGGRAELAAALIPDLDPVRLRLGLE	180
Db	151	VRESSVRAMSAHARVMDGLGLARTPNWLLHGGKGGRAELAAALIPDLDPVRLRLGLE	210
QY	181	NDERAYSPAELLPICATGTPLVFDAAHHVHVHDKLPDQEDPSVREWLRLARATWOPPEWQ	240
Db	211	NDERAYSPAELLPICATGTPLVFDAAHHVHVHDKLPDQEDPSVREWLRLARATWOPPEWQ	270
QY	241	VVHLSNGTFGPQDRRHSHLIADFPSPAYADVPOTEVEAKGKEEAIALRLMAPFK	294
Db	271	VVHLSNGTFGPQDRRHSHLIADFPSPAYADVPOTEVEAKGKEEAIALRLMAPFK	324

RESULT 2
S55418
UV-endonuclease homolog ywjd - Bacillus subtilis
C:Species: Bacillus subtilis


```

Query Match          6.5%; Score 98.5; DB 2; Length 313;
Best Local Similarity 25.1%; Pred. No. 0.64;
Matches 65; Conservative 35; Mismatches 84; Indels 75; Gaps 15;

QY 11 PEVRFRTVLSRYALSP-AEREAKLL--DLYSSNIKTLCGAADYCAAHDIRLYRLSS 66
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 55 PELKGLV-----AASPGPEAARILCVGLHSGVCLRAHPSP--ASLSLLPYR--EY 105
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 67 LPPMLDAGDDTGAVALTHLAPQLLEAGHAFTDAGVRL--MHP-----EQFIVLNSDRPE 120
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 106 EFTLVAFTGDPKDSRV-HLA---EASLLGARGLYVVGPMHPGYBERLSMLCASRVE 160
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 V-RESSVRAMSAHARVMDGLGLARTPWNLLLLHGCKGR----- 158
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 161 APREAPVLSMS-----IASLVTPRL-LGMREGRFRGTEIALSGSTAWLRFERAG 209
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 159 -----GAELAALPDLPDPVRLRLGLENDERAYSPAELLPICEATGTPVLVFAHHHV 211
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 210 EIEEARGSSVEAYSTPLGLPGALVLYLSRQSRAVHVHLEMLPLHRTPTPIVF----- 262
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 212 HDKLPDOEDPSVREWVLA 230
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 263 ---MASVEEPSYRD-VIRS 277
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
F86785
glutamate synthase (NADPH) (EC 1.4.1.13) large chain [imported] - Lactococcus lactis sub
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Mar-2001
C:Accession: F86785
R:Boitoin, A.; Wincker, P.; Mauder, S.; Jaillon, O.; Malarne, K.; Weissbach, J.; Ehrlich
Genome Res. in press, 2001
A:Title: The complete genome sequence of the lactic acid bacterium.
A:Reference number: A86625
A:Accession: F86785
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1486 <STO>
A:Cross-references: GB:AE005176; NID:g12724262; PIDN:AAK05384.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: gltB
C:Superfamily: glutamate synthase (NADPH)
C:Keywords: oxidoreductase

Query Match          6.5%; Score 98.5; DB 2; Length 1486;
Best Local Similarity 23.2%; Pred. No. 5;
Matches 66; Conservative 33; Mismatches 81; Indels 105; Gaps 14;

QY 14 RPRVTLSRYALSPAEREAKLLDLYSSNIKTLCGAADYCAAHDIRLY----- 61
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 233 RFSTNTPSWNRAOPFRFLA-----HNGEINTLRGAENMKVNDIEMYNENSDSAKLEN 287
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 62 -----RLSSSLFPMLDLA-GDDTGAA-----VLTHLAPQLLEAGHAFTD-- 99
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 288 CMVEFYRNGRELPSLLTWIPEAWGEOTGUSPELKAFYEYSTAHIAFPWGPAAALVFTDGK 347
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 100 -AGVRL-----LMHPQFQTVLNSD-----RPEVRESSVRAMSAHARVMDGLGL 141
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 348 TVGARLDRNGRLPSRYLVTKNSIILSSGSGVVDIPADEIIEKSV-----LG- 394
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 142 ARTPNWLLHGGGKG--RGAEALALPD-LPDPVRLRLGI-----ENDERAYSPA- 190
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 395 ---PGNMLLVNTDECKTIRNEVKSYANKYPQEFSLAGLKLKLSALTESEKTNPIPSAK 451
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 191 -----LLPICEATGTPLV---FDAHHHVHDK 214
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 452 MNTLWKLFGYTDVIRIVLPLMAESANEPTISMGFDAPLAVLSQ 496
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 9
C86772
conserved hypothetical protein yljf [imported] - Lactococcus lactis subsp. lactis (st
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
C:Accession: C86772
R:Boitoin, A.; Wincker, P.; Mauder, S.; Jaillon, O.; Malarne, K.; Weissbach, J.; Eh
Genome Res. in press, 2001
A:Title: The complete genome sequence of the lactic acid bacterium.
A:Reference number: A86625
A:Accession: C86772
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-535 <STO>
A:Cross-references: GB:AE005176; NID:g12724145; PIDN:AAK05277.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yljf

```

```

Query Match          6.3%; Score 96; DB 2; Length 535;
Best Local Similarity 24.4%; Pred. No. 2.1;
Matches 68; Conservative 36; Mismatches 87; Indels 88; Gaps 17;

QY 11 PEVRFRTVLSRYALSPAEREAKLLD--LYSSNIKTLCGAADYCAAHDIRL---YRLS 64
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 14 PVSXKVSILVKLRL-----QESDLLDYRIKESIDARHGEIDFIYVDIKLKDEARIL 68
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 65 S-----SLFPMPLDLAGDDTGAVALTHLAPQLLEAGHAFTDAGVRLMHPEQFIVLNSDR 118
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 69 SKRIKVNLSAPELDYVNPVIGSKKMLH-RPVVI---GEGPAG---MFAALLLAQNGYR 119
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 119 PEVRESSVRAMSAHARVMDGLGLARTPWNLLLLHG-----CKGGRG---AELAA 164
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 120 PIVLRG-JAVDERVKSIDKF-----W-----LEGKLNPKSNVQFEGGAGATFSGDKLTS 168
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 165 LIPDLPDPVRLRLGLENDERAYSPAELLPICEATGTPVLVFAHHHVHDKLPD----- 217
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 169 RVKRL-----RGRKVLVEFVQAGAPEDIL-----YKAHPHVGTDLLRDIKVNIRK 213
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 218 -----QEDSVREWVLRARATWQPPWQVHLSNG 247
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 214 QITELGGEVHFDAQVEEFLI-----EDEELQAVKLADG 246
   ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 10
T02583
probable ABC transporter At2g39190 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein T16B24.17
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02583; C84814
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Syke
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.
A:Reference number: Z14679
A:Accession: T02583
A:Status: translated
A:Molecule type: DNA
A:Residues: 1-791 <ROU>
A:Cross-references: EMBL:AC004697; NID:g3402671; PID:g3402685
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: C84814
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-791 <STO>
A:Cross-references: GB:AE002093; NID:g3402685; PIDN:AAK28988.1; GSPDB:GN00139

```

```

C:Genetics:
A:Gene: At2g39190; T1624.17
A:Map position: 2
A:Introns: 110/1; 168/3; 216/1; 233/3; 281/3; 316/3; 358/3; 424/3; 460/2; 514/3; 536/3;

Query Match          6.3%; Score 96; DB 2; Length 791;
Best Local Similarity 24.1%, Pred. No. 3.6;
Matches 83; Conservative 30; Mismatches 107; Indels 124; Gaps 18;

QY 36 LDLYSSNIKTLRGAADYCAAHDIRLYRLSSSLFPMPLDLAGDDTGAAVLTHLAP-----QL 90
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 114 VDLFMAVDRALETYACY-----FYSLSKMWSRRLPEVYDSQAVDYFNCRPHWFAFRL 166

QY 91 LEAGHAFTADGAVRL-----LMHPEQ-----FIVLN---SDRPEV 121
   || ||| ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 167 LEVFSFTAAIRLRTFSASVKGSLEASQIFGVMLKETMLHLGFTFIKVGQSLSLTPDI 226

QY 122 --RESSVRAMSAHARVMDGLGLARTPW--NLLLLHGGKGG----- 157
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 227 IGTEISKALSELHDRI-----PPFPWPEAVKILBSELGGPVEFSFQSQETVAAASFG 280

QY 158 -----RGAELAAI--PDLDPVR-----LRGL-----ENDERAYS----- 187
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 281 QVYRGRTLQDGVAVKVRQPDRLRHAVLRDIYTLRLGLGLVLRKVKAKRENDIRVYADELGMG 340

QY 188 -PAELLPLICEATGCTPLVFDAHHHVHDKLPDOEDPSVREWHVLRARATWOPPEWQVWHLN 246
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 341 LAGELDTLEAANASEFQEAHARFSYIRVPKYVQHILTRKRVITM-----ENWVG---- 389

QY 247 GIEGPODRHSHLIADFPFSAYADVPOIEVEAKGKEETAALRLM 290
   || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 390 --ESPDT-----LLA-ISSGYSD---HONESHEKQKIEARRLL 422

```

```

RESULT 11
T36004
hypothetical protein SCC22.14c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 08-Sep-2000
C:Accession: T36004
R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21574
A:Accession: T36004
A:Status: Preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-368 <SEE>
A:Cross-references: EMBL:AL096839; PIDN:CAB50757.1; GSPDB:GN000070; SCOEDB:SCC22.14c
A:Experimental source: strain A3(2)
C:Genetics:
C:Gene: SCOEDB:SCC22.14c
C:Superfamily: Streptomyces coelicolor hypothetical protein SCC22.14c

```

Query Match	6.2%	Score 95	DB 2	Length 368
Best Local Similarity	26.0%	Pred. No. 1.6		
Matches 81	Conservative 27	Mismatches 103	Indels 100	Gaps 18
Qy	11	PEVRFRTVLSRYRALSPAEREAKLLDLYSSNIKT-----LRCAADYCAAHD--	57	
Db	66	PPARY----LARRRELGVMEADRRLL--LRGSGITTYLVDAGLPGDLTGPTEMATAADT	119	
Qy	58	--IRLYRSLSSLPFMDLAGDDTCAAYLVTHLAPOLLEAGH-----AFTD-AGVR--L	104	
Db	120	REIVRLLEAEQV-----ADTSGTVESFLA-NLAERVHGAANNAVAFTSVACVRHGL	170	
Qy	105	LMPEQFIVLNSDRPVESSVFRMSAHARVWDGGLGLARTPWNLLLLHGKGGKGAELAA	164	
Db	171	ALAPEP-----PGFGEVGRGAAARWLT-----GREYGG	197	
Qy	165	LIPDLDPDVRRLGLENDRAYSPAELLICEATGTPLVFDDAHHHVHDKLP-DQEDPSV	223	

```

Db 198 ---ELSDPVLLR-----HLLWITAVASGLPL--QLHAGLGPGLURIDRTDPLV 239
Qy 224 REWVLRARATWQPPWVHLSNGIEGPQDRRHSHLIADFPISAYADVPQIEVEAKGKEA 283
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 240 LTDVVRTTAGLTGD---LVLLHGY--PYHRHAHLACGVPHVYADSGAALVRTGARAAT 293
Qy 284 IAAELRL-MADFP 293
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 294 VLAIEILELAPF 304

RESULT 12
H75579
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: H75579
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, R.J.;
  M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
  S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: H75579
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-515 <WHI>
A:Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12492.1; PID:g646
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0265
A:Map position: 2

```

```

Query Match      6.2%; Score 95; DB 2; Length 515;
Best Local Similarity 24.5%; Pred.No. 2.5;
Matches 48; Conservative 16; Mismatches 86; Indels 46; Gaps 7;

Qy 78 TGAAVLTILAPOLLEAGHAFTDAGVRLLMHPQEQTIVLNSDRPEVRESSVRAMSAHARYMD 137
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 109 TAPAAARMAVALEPF-----LQRLEPF-----GGDLPEALSGVQTLLTVPYLRQAD 154

Qy 138 GLGLARTETWNLILL-----HGCKGGRGAELAALIPDLDPVRLRLGLGNDEARAYSPA 189
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 155 LLELAGEAWQLWQLARRQPRPAHSAQAGRERVRVVLGGQVPDRVDWNNRTLLDLWGRGEPP- 213

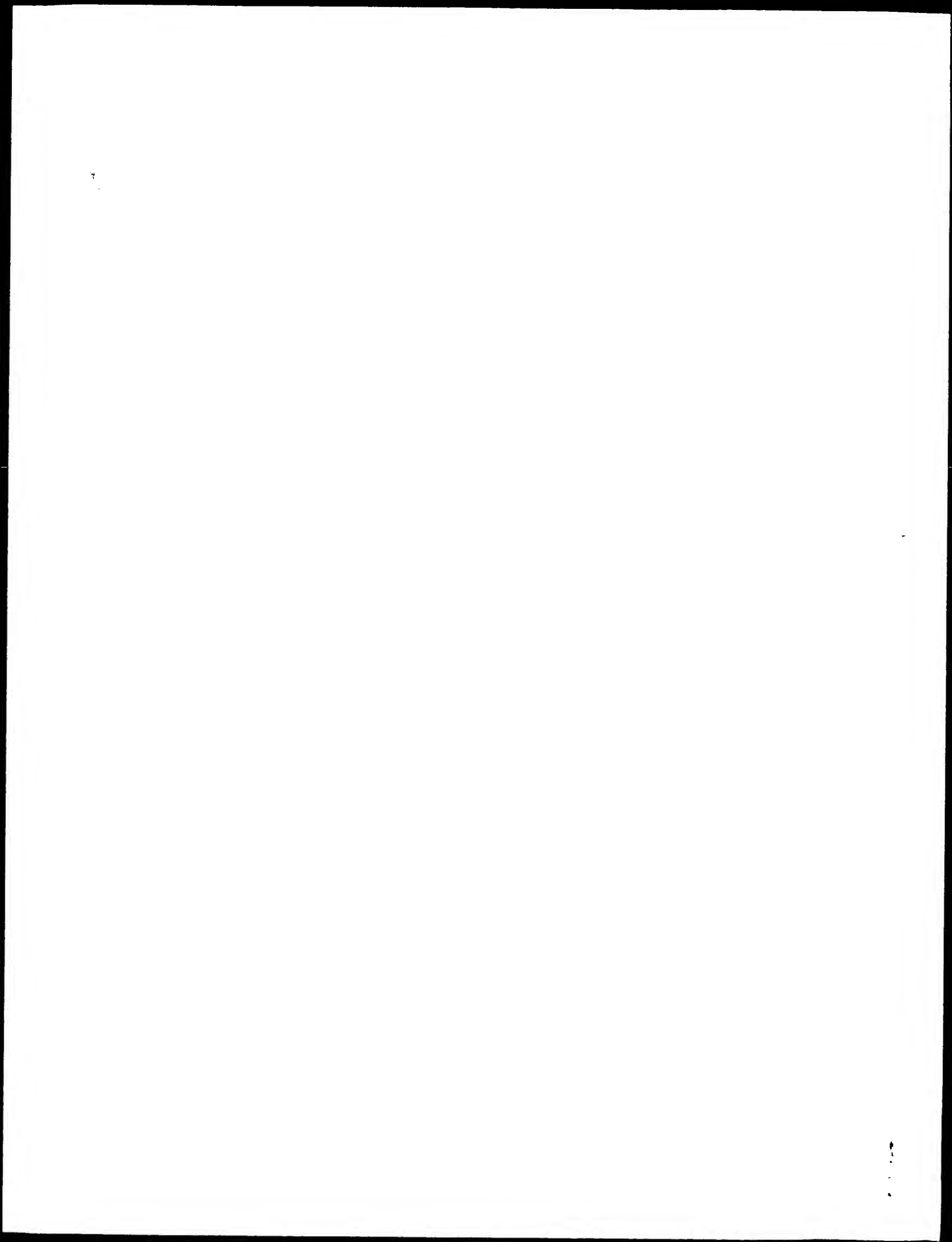
Qy 190 ELLPICEATGTPLYFDAAHHVHHKPLDQEDPSRVENWVLRARATWQPPWQVHLSNGIE 249
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 214 -----HHVALD-LFAPQPPA---LAALRELWQAEAAQLATGPE 251

Qy 250 GPQ-DRRHSHLIADFP 264
      : | | : | | |
Db 252 RDEVORRFARALAEFP 267

```

RESULT 13

t17403
pyochelin synthetase (EC 6.3.2.-) - *Pseudomonas aeruginosa*
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C:Accession: t17403
R:Reimann, C.; Serino, L.; Beyeler, M.; Haas, D.
Microbiology 144, 3135-3148, 1998
A:Title: Dihydroaeruginosic acid synthetase and pyochelin synthetase, products of the
A:Reference number: Z18770; MUID:99061207
A:Accession: t17403
A:Status: preliminary; translated from GB/EMBL/DBD3
A:Molecule type: DNA
A:Residues: 1-1809 <REI>
A:Cross-references: EMBL:AF074705; NID:g3386352; PID:g3386354; PIDN:AAC83657.1
C:Genetics:
A:Gene: pchF
C:Function:




```

DR InterPro: IPR003265; Endo_3c.
DR Pfam: PF00730; HhH-GPD; 1.
DR SMART: SM00391; MBD; 1.
DR Endonuclease
KW SEQUENCE 580 AA; 66050 MW; BF16FB21A34B8E5F CRC64;

Query Match      100.0%; Score 3055; DB 4; Length 580;
Best Local Similarity 100.0%; Pred. No. 1.9e-219;
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGTGLESLSLGRGAAPTIVTSERLVPDPNDLRKEDVAMELERVGEDEQMMIKRSSE 60
DB 1 MGTGLESLSLGRGAAPTIVTSERLVPDPNDLRKEDVAMELERVGEDEQMMIKRSSE 60

QY 61 CNPLQLEPIASQAFCATAGTECRKSVPCGWERVVKORLFGKTAGTGFDFYFISPOGLKFRS 120
DB 61 CNPLQLEPIASQAFCATAGTECRKSVPCGWERVVKORLFGKTAGTGFDFYFISPOGLKFRS 120

QY 121 KSSLANYLHKNGETSLKPEDDFTVLSKRGTKSRKYKDCSMAALTSHLQNSNNWNLR 180
DB 121 KSSLANYLHKNGETSLKPEDDFTVLSKRGTKSRKYKDCSMAALTSHLQNSNNWNLR 180

QY 181 RSKCKKDVMPSPSSSELSQESRGLSNFTSTHLLKDEGVDVNFVRKPKGKVTILKG 240
DB 181 RSKCKKDVMPSPSSSELSQESRGLSNFTSTHLLKDEGVDVNFVRKPKGKVTILKG 240

QY 241 IPKTKTKGCRKSCGFGVOSDSKRESVCNKADAEPVAQKSQLDRVVCISDAGACGETL 300
DB 241 IPKTKTKGCRKSCGFGVOSDSKRESVCNKADAEPVAQKSQLDRVVCISDAGACGETL 300

QY 301 SVTSEENSLVKKERKSLSSGNSFCSEQKTSGLINKFCSEAKDSEHNEKYEDTFLESEELGT 360
DB 301 SVTSEENSLVKKERKSLSSGNSFCSEQKTSGLINKFCSEAKDSEHNEKYEDTFLESEELGT 360

QY 361 KVEVVERKEHLHTDILKRGSEMDNCSPTKDFTEGKIFQEDTIPRTQIERKRTSLYFSS 420
DB 361 KVEVVERKEHLHTDILKRGSEMDNCSPTKDFTEGKIFQEDTIPRTQIERKRTSLYFSS 420

QY 421 KYNKEALSPRRKAPKKTWPPSPENLVQETLFDHPWKLLIATIFLNRTSGKMAIPVLWK 480
DB 421 KYNKEALSPRRKAPKKTWPPSPENLVQETLFDHPWKLLIATIFLNRTSGKMAIPVLWK 480

QY 481 FLEKYPSEAVARTADWROVSELLKPLGLYDLRAKTIKFSDEYLTQKWKYPTELHGIGKY 540
DB 481 FLEKYPSEAVARTADWROVSELLKPLGLYDLRAKTIKFSDEYLTQKWKYPTELHGIGKY 540

QY 541 GNDSYRIFCVNWKQVHPEDHKLKHYDWLWENHEKLSLS 580
DB 541 GNDSYRIFCVNWKQVHPEDHKLKHYDWLWENHEKLSLS 580

RESULT 2
Q92D7 PRELIMINARY; PRT; 554 AA.
AC Q92D7;
DT 01-MAY-1999 (TremBLrel. 10, Created)
DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE METHYL-CPG BINDING PROTEIN MBD4.
GN MBD4.
OS Mus musculus (Mouse), and
OS Mus musculus domesticus (western European house mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090; 10092;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98449942; PubMed=9774669;
RA Hendrich B., Bird A.;
RT "Identification and characterization of a family of mammalian methyl-
CPG binding proteins.";
```

```

RL Mol. Cell. Biol. 18:6538-6547(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=129;
RX MEDLINE=99373255; PubMed=10441743;
RA Hendrich B., Abbott C., McQueen H., Chambers D., Cross S., Bird A.;
RT "Genomic structure and chromosomal mapping of the murine and human
mbd1, mbd2, mbd3, and mbd4 genes.";
RL Mamm. Genome 10:906-912(1999).
DR EMBL: AF072249; AAC68878.1; -.
DR EMBL: AF120996; AAD56595.1; -.
DR MGD: MGI:1333850; Mbd4.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR001739; MBD.
DR InterPro: IPR002052; N6.Mtase.
DR Pfam: PF00730; HhH-GPD; 1.
DR Pfam: PF01429; MBD; 1.
DR SMART: SM00391; MBD; 1.
DR PROSITE: PS00092; N6.MTASE; UNKNOWNL.
SQ SEQUENCE 554 AA; 62577 MW; 792D37CB180291F5 CRC64;

Query Match      59.6%; Score 1821.5; DB 11; Length 554;
Best Local Similarity 66.2%; Pred. No. 1.4e-127;
Matches 384; Conservative 49; Mismatches 116; Indels 31; Gaps 11;

QY 6 LESLSLGD---RGAATVTSERLVPDPNDLRKEDVAMELERVGEDEQMMIKRSSECN 62
DB 1 MESPNGLDNRVGR-----ESLVPDPWDRCKEDIAVGLGVGVDGDKDLVI--SSERS 50

QY 63 PLQLEPIASQAFCATAGTECRKSVPCGWERVVKORLFGKTAGTGFDFYFISPOGLKFRS 122
DB 51 SLQLEPIAST-I-SSTTATEGHKVPFCGWERVVKORLSGKTAGTGFDFYFISPOGLKFRSR 109

QY 123 SLANYLHKNGETSLKPEDDFTVLSKRGTKSRKYKDCSMAALTSHLQNSNNWNLR 182
DB 110 SLANYLHKNGETSLKPEDDFTVLSKRGTKSRKYKDCSMAALTSHLQNETDVSQNKLUKTS 169

QY 183 KCKKDVMPSPSSSELSQESRGLSNFTSTHLLKDEGVDVNFVRKPKGKVTILKG 242
DB 170 KWKTDVLPSPSGTSESPSSGLSNSACLLIREHRDIQDVDEKRRKSKRVTVLKGTA 229

QY 243 IKTKTKGCRKSCGFGVOSDSKRESVCNKADAEPVAQKSQLDRVVCISDAGACGETLSV 302
DB 230 SQTOKCRKSLLESTQRNKRASVVKVQGDRELVPQESQLNRTLCPADACA-RETVGL 288

QY 303 TSEENSLVKKERKSLSSGNSFCSEQKTSGLINKFCSEAKDSEHNEKYEDTFLESEELGT 362
DB 289 AGE-----EKSPGCLDLCFIQVTSGLTNKFSHTEAAGEANR-EQTFLESEELRSK- 338

QY 363 EVVERK--EHLHTDILKRGSEMDNCSPTKDFTEGKIFQEDTIPRTQIERKRTSLYFSS 420
DB 339 --GDRKGEAHLHTGVLDGSEMP-SCSQAKKHFTSE--TFQEDSIPRTQVEKRTSLYFSS 394

QY 421 KYNKEALSPRRKAPKKTWPPSPENLVQETLFDHPWKLLIATIFLNRTSGKMAIPVLWK 480
DB 395 KYNKEALSPRRKAPKKTWPPSPENLVQETLFDHPWKLLIATIFLNRTSGKMAIPVLWE 454

QY 481 FLEKYPSEAVARTADWROVSELLKPLGLYDLRAKTIKFSDEYLTQKWKYPTELHGIGKY 540
DB 455 FLEKYPSEAVARAADWROVSELLKPLGLYDLRAKTIKFSDEYLTQKWKYPTELHGIGKY 514

QY 541 GNDSYRIFCVNWKQVHPEDHKLKHYDWLWENHEKLSLS 580
DB 515 GNDSYRIFCVNWKQVHPEDHKLKHYDWLWENHEKLSLS 554

RESULT 3
Q919FI PRELIMINARY; PRT; 416 AA.
AC Q919FI;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
```


Query Match					
Best Local Similarity		7.2%; Score 220.5;	DB 13;	Length 467;	
Matches		98; Conservative	60; Mismatches	157; Indels	67; Gaps
QY	22	SSERLVPPPPDLRLKEDVAMELER---	VGEDEEOMTKRS--SECNP	LLOEPTASAOFGA	76
Db	49	SSEH-OPEPADEGADMSAEENI,AVP	ESSASPQRORSVIRDGPMDVPD-		99
QY	77	TAGTECRKSVCWGWRVVVKQLFKGT	AGRFVDVFISPOGLKFRSKSLANYL	HKNGETSL	136
Db	100	-----TLPEGWTRKLQKSGSAGKE	DYVLINENGAFKFSVELIAYFOKV	GDTSL	151
QY	137	KPEDFDFTVLKSRGIKSYKDCSMAAL	TSHLQNOSNNNWLRTSKCKKD-V-	FMPSS	194
Db	152	DNPNDFDTV-TGRGPSRREQ-----	KQPKKPAPKASSVSGRGRPKGSIK	KVPPVK	204
QY	195	SSELQESGLSNFTSTHLLK-----	EDEGVDDVNFKV-----RKPGK	VTILKGTP	242
Db	205	SEGVOVKVIEK-SPGKLLVMFYSGW	KEASDATTSQQVLVIKRGGRKKR	KE-TDP	SAAP 262
QY	243	IKKYTKCGRCSCGFVOSDKRESVCN	KADAESPVAQKSOLDRTVCISDA	GACGETLSV	302
Db	263	KKRGKRKSNVSLAAAAEAAKRAI--	-KESSIKPLLE-----TVLP	LIKKTRETIVS	313
QY	303	TSEE-----NSLVKK-----	KERSLSSGSCFNCSEQKTSGLINK	FCSAKDSHENKYE	349
Db	314	DVKDTIRPELTPIVEKMVGONPACS	PESRSTEGSPFKITGLPKKELOQH	HHHHHHHHH	373
QY	350	DTFLESEEIGTKVEVRKEHL 371			
Db	374	HHSESASATSPPETSKONI 395			
RESULT	6				
ID	042403	PRELIMINARY;	PRT;	344 AA.	
AC	042403;				
DT	01-JAN-1998 (TrEMBLrel. 05,	Created)			
DT	01-JUN-1998 (TrEMBLrel. 05,	Last sequence update)			
DE	01-JUN-2001 (TrEMBLrel. 17,	Last annotation update)			
DE	ATTACHMENT REGION BINDING PROTEIN (FRAGMENT).				
GN	ARBP.				
OS	Gallus gallus (Chicken).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Akchosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
OX	Gallus.				
OX	NCBI_taxID=9031;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97415642; PubMed=9271441;				
RT	Weitzel J.M., Buhrmaster H., Straetling W.H.;				
RT	"Chicken MAR-binding protein ARBP is homologous to rat methyl-CpG-				
RT	binding protein MeCP2."				
RL	Mol. Cell. Biol. 17:5656-5666(1997).				
RL	EMBL; Y14166; CAA74577.1; .				
DR	InterPro; IPR000637; AT.hook.				
DR	InterPro; IPR001739; MBD.				
DR	pfam; PF01429; MBD; I.				
DR	PRINTS; PR00929; ATHOOK.				
DR	SMART; SM00391; MBD; 1.				
DR	NON_TER 344 344				
FT	SEQUENCE 344 AA; 33640 MW; 53DD7B9C-GDCE4EF3 CBCE4.				

```

Qy 71 SAQFCATAGTCC-----RKSV-----PCGWVVVVKQRLFGKTAG 104
Db 60 AKGADGGGTAAPAAPEASASPKQRSILDRGPMVDDPTLPEGWTRKLQKRSRGS 119
Qy 105 REDYFIYSPQGLKFRSKSSLIANYLHKNGETSLKPEDFDFTVLKSRGJKSR 154
Db 120 KYDVYLINPQCAKFRSKVELLIAYFEKVGDDSLDNPDFDTV-TGRGSPSR 168

RESULT 7
Q9SF22 PRELIMINARY; PRT; 702 AA.
AC Q9SF22
AC Q9SF22;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE F26K24.5 PROTEIN.
DE F26K24.5
GN Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Konning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsiis thaliana chromosome III BAC F26K24 genomic sequence.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC016795; GenBank; F26K24.5;
SQ SEQUENCE 702 AA: 791713 MW: 6F2EDC7F8ED1F038 CRC64;

```

```

Query Match          6.6%  Score 202.5;  DB 13;  Length 344;
Best Local Similarity 34.7%;  Pred. No.3e-07;
Matches 59;  Conservative 17;  Mismatches 57;  Indels 37;  Gaps 5;

QY 16 AAPTVTSSERLVDPDPDLRKEDYAMELER-----VGDEEQMMIKRSKSECNPLLQEP 70
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5 AAAAAAGGERL-----EQADEGVAGLUKRRPPKAKKGRKERREDPGEAAEPFGSAEAE 59

```

Db 421 EVISRDGHKQTSVFLASIDQRSERAGSACTAL-----VAVIADWFQKNGN 469
QY 503 LKPLGLYDLRAKTIIVKFSDEVLTQWKYPYIELHGIKGYKNDYSYRIFCVNE-WKQVHPEDH 561
Db 470 LMP-----IKSOFDSLIRE-----GSLEWRNLGENETYMOKFPDKH 505
RESULT 8
Q25875
ID Q25875 PRELIMINARY; PRT: 782 AA.
AC Q25875;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE PR86 RHOPTRY PRECURSOR PROTEIN.
GN P82.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HONDURAS I CDC;
RX MEDLINE=90348711; PubMed=2200961;
RA Ridley R.G., Takacs B., Lahm H.W., Delves C.J., Goman M., Certia V.,
Matile H., Woollett G.R., Scaife J.G.;
RT "Characterisation and sequence of a protective rhoptry antigen from
Plasmodium falciparum";
RL Mol. Biochem. Parasitol. 41:125-134 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HONDURAS I CDC;
RX MEDLINE=92244303; PubMed=1574089;
RA Howard R.F.;
RT "The sequence of the p82 rhoptry protein is highly conserved between
two Plasmodium falciparum isolates";
RL Mol. Biochem. Parasitol. 51:327-330 (1992).
DR EMBL; M80807; AAA29717.1;
SQ SEQUENCE 782 AA; 90096 MW; DIAD099862528D42 CRC64;

Query Match 4.7%; Score 143.5; DB 5; Length 782;
Best Local Similarity 21.5%; Pred. No. 0.022;
Matches 112; Conservative 70; Mismatches 175; Indels 165; Gaps 26;
QY 123 SLANYLHKNGET----SLKPEDDF-----TVLSKRGKRSYKDCSMAALTSHLQNS- 171
Db 18 NVADGINVNGDNNYKGTIINDFDNYWYTPINKKEFLNSYED--KFSSESLENKSS 75
QY 172 -NNSNMNLRTRSKCKKDVMPSPSSSELSQESRGLSNFTSTHLLKDEGVDDVNFVRK 230
Db 76 VDDGINLTDSTSNK-----SSKKGHSRVSASAAAILEEDDSKDDMEF----- 122
QY 231 PKGKVTILKGIPIKTKKGRKSGVSDSKRESVNCNKADESEPVAKSOLDRTVCI 290
Db 123 -KASPVVKTSTPSGTQTSGLKSS--PSSTKSSPSN-----V 158
QY 291 SDAGAGETLSVTSENSLVKKERSLSSGSNFCSEQKTSGLINKFCSAKDSHNEKYEY 350
Db 123 -KASPVVKTSTPSGTQTSGLKSS--PSSTKSSPSN-----V 158
QY 291 SDAGAGETLSVTSENSLVKKERSLSSGSNFCSEQKTSGLINKFCSAKDSHNEKYEY 350
Db 159 KSASPHGES---NSSESTTKSKRSAS-----VAGIV-----GADEEAPPAPKN 200
QY 351 TFLESEET-GTKVEVVERKEHLHT-----DILKRGs-----EMDNCSPTKDDFTG 395
Db 201 TLTPLEELYPTNVNLFNFKYSLNNEENINILKNEGDLVAQKEFEYDENMEKAKODKK 260
QY 396 --EKT-----FQEDTTPRTOIERRTKTSLYFSKYKNEALSPPRRKAFKWTTPRS 443
Db 261 ALEKIGKESDEAPFMFSKFNLENQVKNVAGSFRFSK--LNP-----FKK----- 307
QY 444 PFNLVQETLFDHPWKLLIATIFLNTSGKMAIPVLWKFLKYPSSA-----488
Db 308 -DEVIEKT---EVSKKTFSGIGFNLTKKAKVGLGVGTYQYPTMYLNCNPNNSNLFDTI 363
QY 489 -----EVARTADWRDVSSELLKPLGLYDLRAK--TIVKFSDEVLTQWKYPYIEL 534
Db 364 ESLOGRIIDIKKESMISTTFEOKKCLKNMGVLDLELNDTQCKFGT-----410
QY 535 HGIGKYGNDYSYRIF-CVNEWKQVHP-----EDHKLNKYH 567
Db 411 -CIGSFGHEHLRLYEENDLFKPHPNIDYLTADGYKLOKNH 451

RESULT 10
Q26007

QY 489 -----EVARTADWRDVSSELLKPLGLYDLRAK--TIVKFSDEVLTQWKYPYIEL 534
Db 364 ESLOGRIIDIKKESMISTTFEOKKCLKNMGVLDLELNDTQCKFGT-----410
QY 535 HGIGKYGNDYSYRIF-CVNEWKQVHP-----EDHKLNKYH 567
Db 411 -CIGSFGHEHLRLYEENDLFKPHPNIDYLTADGYKLOKNH 451
RESULT 9
Q25730
ID Q25730 PRELIMINARY; PRT: 782 AA.
AC Q25730;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE RHOPTRY ASSOCIATED PROTEIN-1.
GN RAP-1.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FC27;
RA Shi Y., Lal A.A.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20985; AAA63681.1;
SQ SEQUENCE 782 AA; 90023 MW; F69E26A2A564C8EA CRC64;

Query Match 4.6%; Score 141.5; DB 5; Length 782;
Best Local Similarity 21.1%; Pred. No. 0.03;
Matches 110; Conservative 71; Mismatches 176; Indels 165; Gaps 25;
QY 123 SLANYLHKNGET----SLKPEDDF-----TVLSKRGKRSYKDCSMAALTSHLQNS- 171
Db 18 NVADGINVNGDNNYKGTIINDFDNYWYTPINKKEFLNSYED--EFSSESLENKSS 75
QY 172 -NNSNMNLRTRSKCKKDVMPSPSSSELSQESRGLSNFTSTHLLKDEGVDDVNFVRK 230
Db 76 VDDGINLTDSTSNK-----SSKKGHSRVSASAAAILEEDDSKDDMEF----- 122
QY 231 PKGKVTILKGIPIKTKKGRKSGVSDSKRESVNCNKADESEPVAKSOLDRTVCI 290
Db 123 -KASPVVKTSTPSGTQTSGLKSS--PSSTKSSPSN-----V 158
QY 291 SDAGAGETLSVTSENSLVKKERSLSSGSNFCSEQKTSGLINKFCSAKDSHNEKYEY 350
Db 159 KSASPHGES---NSSESTTKSKRSAS-----VAGIV-----GADEEAPPAPKN 200
QY 351 TFLESEET-GTKVEVVERKEHLHT-----DILKRGs-----EMDNCSPTKDDFTG 395
Db 201 TLTPLEELYPTNVNLFNFKYSLNNEENINILKNEGDLVAQKEFEYDENMEKAKODKK 260
QY 396 --EKT-----FQEDTTPRTOIERRTKTSLYFSKYKNEALSPPRRKAFKWTTPRS 443
Db 261 ALEKIGKESDEAPFMFSKFNLENQVKNVAGSFRFSK--LNP-----FKK----- 307
QY 444 PFNLVQETLFDHPWKLLIATIFLNTSGKMAIPVLWKFLKYPSSA-----488
Db 308 -DEVIEKT---EVSKKTFSGIGFNLTKKAKVGLGVGTYQYPTMYLNCNPNNSNLFDTI 363
QY 489 -----EVARTADWRDVSSELLKPLGLYDLRAK--TIVKFSDEVLTQWKYPYIEL 534
Db 364 ESLOGRIIDIKKESMISTTFEOKKCLKNMGVLDLELNDTQCKFGT-----410
QY 535 HGIGKYGNDYSYRIF-CVNEWKQVHP-----EDHKLNKYH 567
Db 411 -CIGSFGHEHLRLYEENDLFKPHPNIDYLTADGYKLOKNH 451


```

QY 285 DRTVCISDAGACGETSVTSEENSLVKKERSLSGSNFCSEBOKTSGIINKFCSAKDSEH 344
Db 155 PSNV--KTASPHGES---NSSESTSKSKRSAS-----VSGIV-----GADEEV 194
QY 345 NEKYEDTELESEET-GTKVEVVERKEHLHT-----DILKRGSEM-----DNNCSPT 389
Db 195 PPAPKNTLTPLEELYPTNVLNFKYSLNNMEENIILKNEGOLVAQKEEFYDENWEKA 254
QY 390 R--KDTGEKI-----FOEDTPTQIERRKTSLSYFSKYNKEALSPPRKAFFK 437
Db 255 KEEKKALEKIGKETDEEPPMETDDKFLNOKVERNVAGFSRFTSK--LNP-----PKK 307
QY 438 WTPRSPENLVQETLPHDPKWLIIATIFLNRTSGKMAIPVLNKFLEKYP-----486
Db 308 -----DEVIEKT---EVSKKTYSIGFNVNDKEAKILGVGATYOEPETMLNCPNNS 357
QY 487 -----SAEVARTAD-----WRDVSSELLKPLGLYDLRAK--TIVKFSDEVLTKQW 528
Db 358 HLPDTISSLQGRULIDIKKRENMISTTFEQOKECLNMGMVLDLELKDTECKFGT-----410
QY 529 KYPIELHGKGYGNDYRYP-CVNEWKQVHP-----EDHKLNYH 567
Db 411 -----CIGSGEHLRLRYEFENDLFKHPNIDYLTLLADGYKLOKNH 451

RESULT 15
QY00G6 PRELIMINARY; PRT: 2209 AA.
AC Q900G6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE VAR, MALAP2.56.
GN VAR.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=36329;
RX [1]
RC SEQUENCE FROM N.A.
RP STRAIN=3D7.
RA Devlin K., Bowman S., Churcher C., Harris B., Harris D., Lawson D.,
RA Quail M., Barrell B.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035475; CAB62897.1;
SQ SEQUENCE 2209 AA; 250779 MW; 7A349F6FEEF9F7CB CRC64;

Query Match 4.5%; Score 137.5; DB 5; Length 2209;
Best Local Similarity 18.8%; Pred. No. 0.24;
Matches 136; Conservative 97; Mismatches 226; Indels 263; Gaps 37;

QY 8 SLSLGRGAAPTPTSERLVPDPNDLRKEDVAME-----LERVGED---EEQM 53
Db 1176 SVTIGGEAGGLOST-----DSKDAARGEXTPLDSDFIKRPYPFRYLEEWGQNFCKERRK 1230
QY 54 MIKR--SSECNPLLOPIASAOFGATAGTECRKSPV-----CG-----W- 90
Db 1231 RLKDIKYEORG--DENI--TRYGSGYGEDCKNNLPENPSTFKDLEYPTCAKYCRFYKKWI 1286
QY 91 -----ERVVKORLEFGTAGRFDYFTISPOGLKFRSKSSLYNLHK-----NGE 133
Db 1287 NTKKTEYEKQEKIYVQOQKDATSDNGKNYDSNCDG-KLKQYASTESFLEKLVQCKDNGE 1345
QY 134 TSLK---PEDFDFTVLSKRGIKRYKDCSMAALTSHLQNSNNSNWLRTSKCKKD---187
Db 1346 GTIKFNGGQTQHT-----EDCKSCS-----KFRKICDNDKCS 1378
QY 188 -----VFMPSPSSSLOPESGLSNFTSTHLLKEDGV-----DDVN-----FRKVR 229
Db 1379 GGNTRVKDGKTPIDAKEIANNINSPOEVTMLVSDNGATGFKGDDLKEACEGKGFESIR 1438
QY 230 KPGKVTILKGIPIKTKKCKRSCSGFVQSD-----SKRE 265

```

```

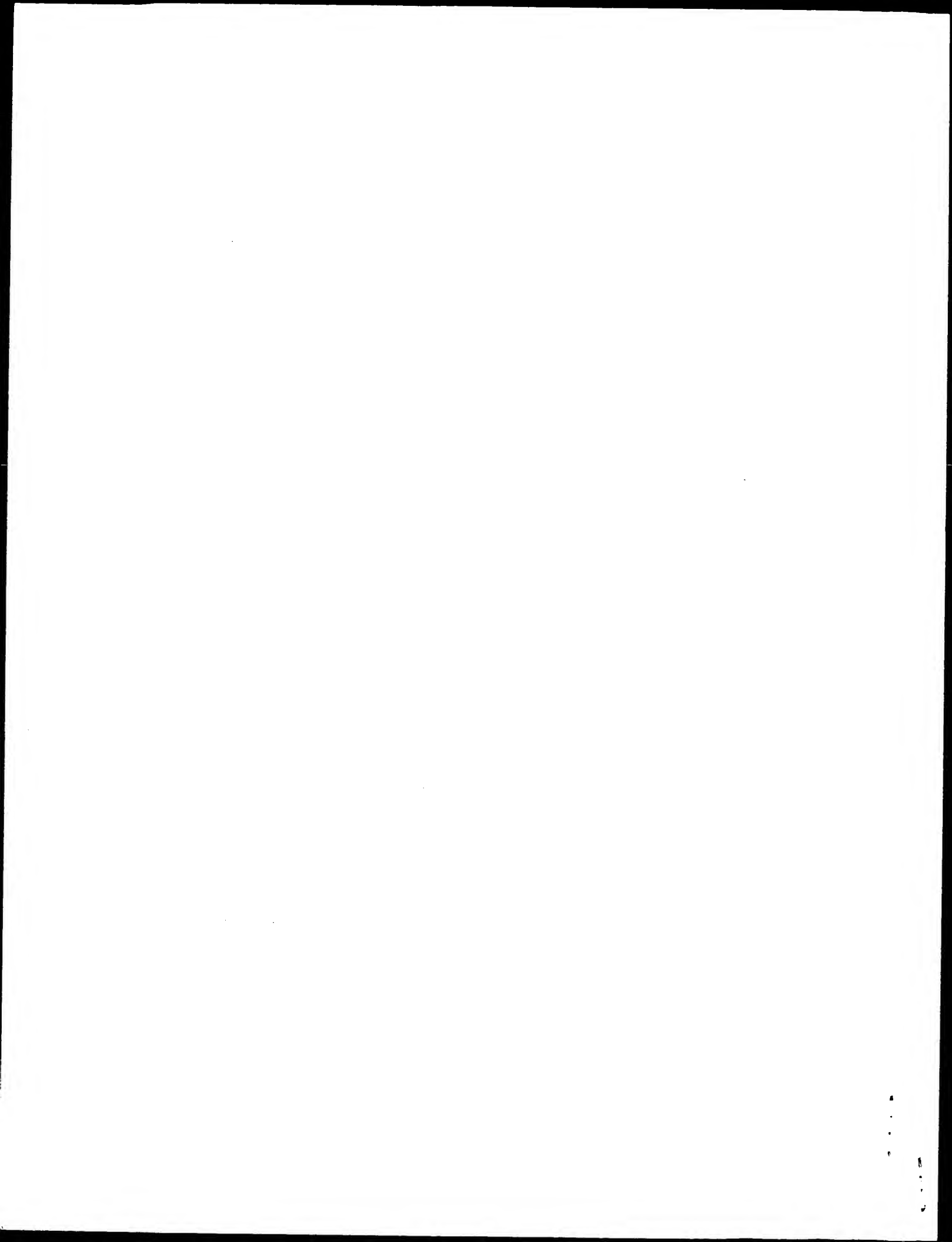
Db 1439 KEQWKCDNVGYVVCCKPKEGRNRETVRGKNDKHHIITIRALVTHWQVFLDDYKKIKHKI 1498
QY 266 SVCNKADAESFVAQKSOLDRTVCISD-----292
Db 1499 SHCTKTDOGS--TCQNKCONCKCVCGEWIPKKREWQOIKDRFLKQYKNKLDLDEDFNLRS 1556
QY 293 -----AGA-----CGETLSVTSEENS-----LVKKERS 316
Db 1557 CLETFVLQIGAAYGEDKFKVILKLSVFDQSCGSAIASSQKNGEYKDAIECMCLKLEEK 1616
QY 317 LSSGSNFCSEOKTSG-IINKFCSAKDSEHNEKYEDTFLSEEEIGTKVEVVERKEHLHTDI 375
Db 1617 -----ANKCKEDHSSGEOTEKECQSPSVEDE--DDTLHEETEVE-----KAPEICKDV 1662
QY 376 LKRGSEMDNN--CSP--TRKDF-----TGEKIFQEDTIPRTOIERRKTSLYFSSKYNKE 425
Db 1663 IKAPTEPEERKACDPAPPTPKETSPATDSKGETNTEPYTFQDQSDPTKTPKEGPKPPKS 1722
QY 426 ALSPPRKAFAKKWTPPRSPFNLVQETLPHDPKWLIIATIFLNRTSGKMAIPVLNKFLEKY 485
Db 1723 LPQPPROK--REFTP-----SDWKKVNSASAF-PWTVGVAFWALSVWVMKKK 1766
QY 486 PSAEVARTADKROVSEL-LKPLGLYDLRAKT--LVKESDEYLTQWKYPIELHIGIKGYKN 542
Db 1767 SKPRV-----DLFSVMEIPQNDYGMPTLKSNNRYVPYSSGKYRGKTYLY-VE-----GDSGT 1817
QY 543 DS 544
Db 1818 DS 1819

```

Search completed: January 15, 2002, 14:06:31
Job time: 970 sec

us-09-724-296-38.rspt

Tue Jan 15 14:12:18 2002



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:52:31 ; Search time 74.94 Seconds
(without alignments)
322.222 Million cell updates/sec

Title: US-09-724-296-37
Perfect score: 1662
Sequence: 1 MIFRFGVSNAMSLWDASPA.....ELSSIRGVKRTGGALQWKS 317

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 68: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1650.5	99.3	320	2 S55418	UV-endonuclease ho
2	484.5	29.2	322	2 E83974	hypothetical prote
3	359.5	21.6	599	2 S71134	UV-endonuclease -
4	345	20.8	656	2 S55262	UV-endonuclease -
5	293.5	17.7	326	2 C75350	probable UV damage
6	104.5	6.3	315	2 B59093	hypothetical prote
7	101	6.1	283	2 S74063	hypothetical prote
8	100	6.0	823	2 G83905	hypothetical prote
9	98	5.9	331	2 S54263	rep A protein - Ba
10	96.5	5.8	969	2 T38478	RhoGAP/LIM domain
11	95	5.7	775	2 T22200	hypothetical prote
12	94.5	5.7	553	2 T21233	hypothetical prote
13	94.5	5.7	1145	2 A59251	myosin - Acetabula
14	93.5	5.6	805	2 T32377	hypothetical prote
15	93.5	5.6	1036	2 T1565	probable isoleucin
16	93	5.6	674	2 T22733	hypothetical prote
17	92.5	5.6	601	2 T02633	hypothetical prote
18	92	5.5	428	2 G75064	hypothetical prote
19	91.5	5.5	921	2 C81153	type I restriction
20	91.5	5.5	237	2 B82644	5-amino-6-(5-phosp
21	91.5	5.5	522	2 G02522	sorting nexin 1 -
22	91.5	5.5	1409	1 ORFFCP	copla polyprotein
23	91.5	5.5	2334	2 S32920	cell wall-associat
24	91	5.5	558	2 T48150	stress-induced pro
25	91	5.5	852	2 B84001	ribonucleoside-dip
26	91	5.5	1020	2 E86165	hypothetical prote
27	91	5.5	1284	1 WNVZAI	A-type inclusion p
28	90.5	5.4	233	1 R5HGIT	ribosomal protein
29	90.5	5.4	279	2 T15313	hypothetical prote

30	90.5	5.4	448	2 T15589	hypothetical prote
31	90.5	5.4	496	1 S41192	cytochrome P450 4D
32	90.5	5.4	625	2 T32739	hypothetical prote
33	90.5	5.4	629	2 E84610	outer membrane pro
34	90	5.4	572	1 F0HYJH	retrovirus-related
35	90	5.4	848	2 C82712	ATP-dependent seri
36	90	5.4	2104	2 T38774	myosin-3 heavy cha
37	89.5	5.4	279	2 T16816	hypothetical prote
38	89.5	5.4	429	2 S29565	apolipoprotein A-I
39	89.5	5.4	553	2 T24639	hypothetical prote
40	89.5	5.4	5255	2 T31677	bacitracin synthet
41	89	5.4	489	2 S77357	aspartate-LRNA li
42	89	5.4	550	2 I64203	type I site-specif
43	89	5.4	1033	2 S02168	mycosubtilin synth
44	89	5.4	5369	2 T44807	Ig light chain-bin
45	88.5	5.3	719	2 A42808	

ALIGNMENTS

RESULT 1

S55418 UV-endonuclease homolog ywjd - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 15-Oct-1999
C:Accession: S55418; B70060

R:Glaser, P.; Danchin, A.

Submitted to the EMBL Data Library, May 1995

A:Description: Cloning and sequencing of the Bacillus subtilis chromosomal region fro

A:Reference number: S55414

A:Accession: S55418

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-320 <GUA>

A:Cross-references: EMBL:Z49782; NID:g853752; PIDN:CAA89865.1; PID:g853757

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivoita, C.; Roche, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili

A:Reference number: A69580; MUID:98044033

A:Accession: B70060

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-320 <KUN>

A:Cross-references: GB:299123; GB:AL009126; NID:g2636240; PIDN:CAB15748.1; PID:ell862

A:Experimental source: strain 168

C:Genetics:

A:Gene: ywjd

Query Match 99.3%; Score 1650.5; DB 2; Length 320;
Best Local Similarity 99.1%; Pred. No. 1.2e-128;
Matches 317; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

Qy 1 MIFRFGVSNAMSLWDASPAKTLTFARYSKLSKTEREALLTVTKANLRMTMTLHVILG 60

Db 1 MIFRFGVSNAMSLWDASPAKTLTFARYSKLSKTEREALLTVTKANLRMTMTLHVILG 60

Qy 61 HGIPLYRFSSSIVPLATHPDVMDVTFPFQKEFGEIGELVKTHQLRTSFHPNFTLTSP 120

Db 61 HGIPLYRFSSSIVPLATHPDVMDVTFPFQKEFGEIGELVKTHQLRTSFHPNFTLTSP 120

121 KESVTKNAVTDMAHYRMLAMGIADRSVINIHIGGAYGNKDTATATQFHNKIQLPQEI 180
121 KESVTKNAVTDMAHYRMLAMGIADRSVINIHIGGAYGNKDTATATQFHNKIQLPQEI 180
181 ERMTLENDKTYTTEETLOVCEQEDVPFDFHHFYANPDHDLNVALPRMUKTWERIG 240
181 ERMTLENDKTYTTEETLOVCEQEDVPFDFHHFYANPDHDLNVALPRMUKTWERIG 240
241 LOPKVHLSPPKSEQAIRSHADYVDANF---LLERFRWGTNIDFMIEAKQKALLRLMD 297
241 LOPKVHLSPPKSEQAIRSHADYVDANFLLERFRWGTNIDFMIEAKQKALLRLMD 300
298 ELSSIRGVRKRGCGALQWKS 317
301 ELSSIRGVRKRGCGALQWKS 320

RESULT 2
E83974
hypothetical protein BH2597 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: E83974
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20263314
A:Accession: E83974
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-322 <STO>
A:Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BAB06316.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2597

Query Match 29.2%; Score 484.5; DB 2; Length 322;
Best Local Similarity 32.2%; Pred. No. 1.9e-32;
Matches 106; Conservative 67; Mismatches 133; Indels 23; Gaps 6;

QY 1 MIPRFGVSNAMSLDASPAKTLTPARYSKLSKTE---RKEALLTVTKANLRNMTLHY 57
DB 1 MRIQGYVAMSMELANSPSKTMTATQEKIEDHAGLRK--LERIAKTNLHNCRLRLK 58
QY 58 ITHGIPLYRFSSSIVPLATHP-DVMWDFVTPFOKEFEIGELVKTHLRTSFHPNFTL 116
DB 59 NLAYQLSFRSLSKLVPLNHPHPLTEGWKYEIAEELQAVGEFASHEQMRIDFHPDHEV 118
QY 117 FTSPKESVTKNAVDMAYHYRMLAMGIADRSVINIHIGGAYGNKDTATATQFHNKIQLP 176
DB 119 LNSEAKEITRSLQTLVHYKLLKGMEDPRHRCVHLVGGKKKGVEAGLEQFIENTASIP 178
QY 177 OEIKERMTLENDKTYTTEETLOVCEQEDVPFDFHHFYANPDH-----ADLNVALP 230
DB 179 KSLLSMIMLENDKSTYIDVLYLGEKLAIPVLDIHH-----HDVLRKSKSLQETWQ 231
QY 231 RMKWTWELGLOPKVHLSPPKSEQAIRSHADYVDANFLLERFRWGTNIDFMIEAK 286
DB 232 RIVATWEDSPLPVKHLSSPLSGEDDPRHHDYINADRTAFLEHIGADAVDHLHVMIEAK 291
QY 287 OKDKALLRLMDLSSIRGVRKRGCGALQW 315
DB 292 KKDALLFQMLKDLAEYITVVSKEAVEF 320

RESULT 3
S71134
UV-endonuclease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 20-Jun-2000
C:Accession: S71134; T39815

R:Takao, M.; Yonemasu, R.; Yamamoto, K.; Yasui, A.
Nucleic Acids Res. 24, 1267-1271, 1996
A:Title: Characterization of a UV endonuclease gene from the fission yeast Schizosacch
A:Reference number: S71134; MUID:96188860
A:Accession: S71134
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-599 <TA>
A:Cross-references: EMBL:D78571; NID:g1399000; PIDN:BA11415.1; PID:g1399001
A:Experimental source: strain SP972
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Lucas, M.; Gaillardin, C.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z21881
A:Accession: T39815
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-599 <LYN>
A:Cross-references: EMBL:AL023859; PIDN:CAAL9577.1; GSPDB:GN00067; SPDB:SPBC19C7.09c
A:Experimental source: strain 972h; cosmid c19C7
C:Genetics:
A:Gene: UVDE; SPBC19C7.09c
A:Map position: 2

Query Match 21.6%; Score 359.5; DB 2; Length 599;
Best Local Similarity 29.7%; Pred. No. 1e-21;
Matches 83; Conservative 60; Mismatches 111; Indels 25; Gaps 7;

QY 34 TERKEALLTVTKANLRNMT---RTLHYIIGHGIPLYRFSSSIVPLATHPDMWDFVTPP- 89
DB 276 TTDGLESVKQGLGTQNVLDLKLVEWNNINFGIHPRVSSDLFPFASH--AKYGYTLFPA 333
QY 90 QKEFREIGELVKTHLRTSFHPNFTLFTSPKESVTKNAVDMAYHYRMLAMGIADR-- 147
DB 334 QSHLEEVGLANKYNHRLTMHPQYQTQIASPREVVVDASIRDLAYHDEILSRMKLNEQLN 393
QY 148 --SVINITHIGGAYCNKDTATATQFHNKIQLPQEIKERMTLENDKTYTTEETLOVCEQED 205
DB 394 KDAVLIIHIGGTPGEGKRETLDRFRKNYQRLSDSKARLVLENDVSVQDILLPQCEQLN 453
QY 206 VPFVDFHFFYANP---DDHADLNVALPRMKTWELGLOPKVHLSPPKSEQAIF----- 256
DB 454 IPLVLDHNNIYVPGTLREGSLDMLPLIPTIRETWKGTQKOHYSASADPTAISGMKR 513
QY 257 RSHADYVDANFLLERFRWGTNIDFMIEAKQKDKALLRL 295
DB 514 RAHSDRV-----FDPPPCDPTMDLMEAKEKEQAVFEL 546

RESULT 4
S55262
UV-endonuclease - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
C:Accession: S55262
R:Yajima, H.; Takao, M.; Yasuhira, S.; Zhao, J.H.; Ishii, C.; Inoue, H.; Yasui, A.
EMBO J. 14, 2393-2399, 1995
A:Title: A eukaryotic gene encoding an endonuclease that specifically repairs DNA dam
A:Reference number: S55262; MUID:95292980
A:Accession: S55262
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-656 <YAJ>
A:Cross-references: EMBL:D11392; NID:g1526560; PID:g927215
C:Genetics:
A:Gene: uvel

Query Match 20.8%; Score 345; DB 2; Length 656;
Best Local Similarity 32.3%; Pred. No. 1.8e-20;
Matches 87; Conservative 51; Mismatches 105; Indels 26; Gaps 8;

QY 46 ANLRNMTLHYIIGHGIPLYRFSSSIVPLATHPDMWDFVTPFOKE-FREIGELVKTHQ 104

Db 277 ANARDIVKMLCWNEKYGIRLRUSSEMFPPASHP-VHGYYKLAPFASEVLAEAGRVAELG 335
QY 105 LRTSFHPNQFTLTSKPESVTKNAVTDMAHYHRLMEAMGIADR---SVNIHHIGGAYGN 160
Db 336 HRLTTHPGQFTQLGPSRKEVESAIRDLEYHDELLSLKLPEQONRDVAVIHHMGOFCD 395
QY 161 KOTATAFOHNIKOLPOEIKERTWLTENDDKTYITTEFLQWCQEDVPFVFDFHHFYANPD 220
Db 396 KAATLFKRNYARLSQSCNRNLVDNDVGWTVHDLLPVCEELNIPMLVDYHHNHICFD 455
QY 221 D----HADLANALP-----RMKWTWRIGLPKVHLGSPKSEQAI-----BSHADYVDANF 267
Db 456 PAHLREGTLDISDPKLOERLANTWKRGIKOKWHYSEP-CDGAVTPPHRRKKRPV---- 510
QY 268 ELERFRGWGTNIDFMIEAKOKKALLRLM 296
Db 511 --MTLPPCPMDLMIEAKDKEQAVFELM 537

RESULT 5
C75350
probable UV damage endonuclease - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: C75350
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, R.J.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: C75350
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-326 <WHI>
A:Cross-references: GB:AE002022; GB:AE000513; MID:g6459590; PIDN:AAF1370.1; PID:g6459590
A:Experimental source: strain R1
C:Genetics:
A:Gene: Dr1819
A:Map position: 1

C:Accession: B59093
R:Okimaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehn J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pX01, the large *Bacillus anthracis* plasmid harboring a *recA* homologue
A:Reference number: A59091; MUID:99445483
A:Accession: B59093
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-315 <OKI>
A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD3322.1; PID:g4894234
A:Experimental source: strain Sterne
A:Note: similar to integrase-recombinase protein (311 aa), MTH893; *Methanobacterium thermophilum* strain 2626
C:Genetics:
A:Gene: pX01-18
A:Genome: plasmid

	Query Match	6.3%	Score 104.5;	DB 2;	Length 315;
	Best Local Similarity	22.8%;	Pred. No. 0.42;		
	Matches	65;	Conservative 41;	Mismatches 114;	Indels 65; Gaps 10;
Qy	32	SKTERKALLTVTKANLRNMRTRLIYIIGIGIPIYRFSSIVPLATHPDVWDFVTFQK	91		
Db	67	SELEHKRYKSQKTRHRIWVVLKLIHTGLGVISPLDGLHLSLIPDQSLSDS--DFITELEE	124		
Qy	92	EFREIGELVAKTHQLRTSFHPNQFTLTSPRESVTKNAVTDMAHYHMLEAMGIAIDRSVIN	151		
Db	125	-----TLKKOTVLST-----KGLTEKQAKYRPL-----IMDRACI	155		
Qy	152	IHIGGAVG-----NKDTATAQFHNKIQLPQEIK-----ERMTLENDKTY-----TTEE	196		
Db	156	INLIWNGLSLQELVSLNMSHIQFARNLIVPGENKLTFRSVSLTWTEDTQQLKYKYYTTIPE	215		
Qy	197	TLQCEQEDVPF--VFDFHH-----FYANPDHADLNVALPRMIKTRWERICLPQKVHLSS	249		
Db	216	PVRQQTNNPLVFAFDNRCGTFERWVYENDAPKALSEVAIQKMIR-----LEV	263		
Qy	250	PKSEQAIRSHADYVDANI'LLERFFROWGTNIDFMTEAKOKKALLR	294		
Db	264	KRAELDRRISACQMENTFILRIEGVTEKGLVSMMGFKTKISLK	308		

```

RESULT      7
S74063
hypothetical protein c0132 - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 09-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 20-Jun-2000
C:Accession: S74063
R:Sensen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan, C.C.Y.; Liu, Q.Y.; Penny,
Mol. Microbiol. 22, 175-191, 1996
A:Title: Organizational characteristics and information content of an archaeal genome
A:Reference number: S73076; MUID:97055432
A:Accession: S74063
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-283 <SEN>
A:Cross-references: EMBL:Y08256; NID:gl707679; PIDN:CAA69434.1; PID:gi707762
A:Experimental source: strain P2
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1999
C:Genetics:
A:Start Codon: TTG
C:Superfamily: Methanobacterium thermoautotrophicum endonuclease IV

Query Match          6.1%; Score 101; DB 2; Length 283;
Best Local Similarity 21.6%; Pred. No. 0.7;
Matches 52; Conservative 48; Mismatches 87; Indels 54; Gaps 13;

Qy    95  ETGLVKTQHQLTSFH-PNQFTLTSPKESV--TKNAVTDMAHYRMLEAMGIADRSVIN 151
       :|::: |: | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    59  EFGVGARELVGVLSVHAFYFNLLSEEREKEIASKQRILDTADRAMELGADAIA----- 112
                                           :|::: |: | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy   152 IHIGGAYGNKDATVAHQHNTIKLPQEIKERM-----TLENDDKTVTTETL 198

```

Db 113 IHI--AFYGMT-PEECYQVKEGLEEVIDKAREMGINVAFGVETMAKETAFGTLEVI 169
 QY 199 QVCEQED--VPVFFHFHYANPDHDLNVALPRMKTWTWIRIGLQPKVHLSPPKSOAI 256
 Db 170 SISKELGVIPYI-DWAHTFARQGEIDYKIIIDRLIK--ELGL--THINS-HFESLV 221
 QY 257 RSHADYVD-----ANFLERFRQWGTNIDFMIEAKOKDKALLRMDLSS 301
 Db 222 YRGGYVDEHIPIDANAPFPELAKELLK-----DISITLICESPELERDALMKWEVL 277
 QY 302 I 302
 Db 278 L 278

RESULT 8
 G83905
 Hypothetical protein BH2047 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650; MUID:20263314
 A:Accession: G83905
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-823 <STO>
 A:Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BA05766.1; GSPDB:GNOC
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH2047

Query Match 6.0%; Score 100; DB 2; Length 823;
 Best Local Similarity 23.18; Pred. No. 4.1;
 Matches 75; Conservative 44; Mismatches 101; Indels 104; Gaps 18;

QY 30 KLSKTERK-----FALLTVTKANLRNMTLHYIIGHIPLYPRESSIVPLATHPDV-MW 83
 Db 332 KRLLEKKRLILIEPLISITEPDSNLERS--WILK-----IKDVEHPDLSLW 377
 QY 84 DFVFPQKEFEIGELVKTHQRTSFH--PNOFTLTPSPKESVTKNVATDMAYHMLEA 141
 Db 378 QERLEWLVKQVAYPYMTLGOKQLFKLESTFTWTGEQP-----RRLG 423
 QY 142 MGIADRSVI-----NIHIGAYGN-----KDTATAQFH--ONIKO---LPQEI 180
 Db 424 EYADRNLLEYECHPLNNIKIGSIKIVLKQDYPKWLSTCAKGEORRQEQALQAEIF 483
 QY 181 ERMTLENDKTYTTEETLOVCEQEDVPFDFHFFHYANPDHDLNVALPRMKTWTWIRIG 240
 Db 484 KMW-----YPNEDS-----VPFLKFVHDLNHPDVH-----TWKRW 515
 QY 241 LQPKVHL-----SSPKSEQAIRSHADYVDANFLERFRQWGTNIDFMIEAKOKDKA 291
 Db 516 QSIKTEIEATHEVASNPKS--VHLSLDYQD-----FDRDMAWLSPDLMI-AKKDDDS 568
 QY 292 LLRLMDELSSIRGVKRGGGALOW 315
 Db 569 YQVILGE---IHDPIVMGWALQF 589

RESULT 9
 S54263
 rep A protein - Bacteroides fragilis
 C:Species: Bacteroides fragilis
 C:Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 08-Oct-1999
 R:Haggoud, A.; Trinh, S.; Mohieddine, M.; Reysset, G.
 submitted to the EMBL Data Library, April 1995

A:Description: Genetic analysis of the minimal replicon of plasmid pIP417 and compari
 A:Reference number: S54263
 A:Accession: S54263
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-331 <HAG>
 A:Cross-references: EMBL:X86702; NID:g804960; PIDN:CAA60390.1; PID:g804961

Query Match 5.9%; Score 98; DB 2; Length 331;
 Best Local Similarity 19.9%; Pred. No. 1.6;
 Matches 52; Conservative 42; Mismatches 85; Indels 82; Gaps 12;

QY 63 IPLYRESSIVPLATHPDVMDVFTPPQKEFEIGELVKTHQRTSFHFNQFTLTSPKE 122
 Db 109 LPYIRKSSVISFQVHPRV-WDCILNFSKGYRKY-ELKAAMFKSQFSMRFYELLSNQK- 165
 QY 123 SVTKNAVTDIMAYHMLEAM-GIADRSVINIHIGGAYGNKDTATATQFHQNIKOLPQEIKE 181
 Db 166 -----TPLYSIDOLKEMFCVADK-----YKRVNDEMFKVVE 197
 QY 182 RMTLENDKTYTTEETLOVCEQEDVPFDFH-----HFY-ANPDHDLNVAL 229
 Db 198 AAKKELDEVS-----PYTFEFTPLKSGRKITAKEYPVYQPEHRDPL-- 240
 QY 230 PRMIKTWIRIGLQPKVHLS--SPKSEQAIRSHADYVDANFLERFRQWGTNIDFMIEAK 286
 Db 241 -----EKHDLOKQVLSWSLSSEVRSVLKNSIGYSD-----KEIKNNLDLFISAQ 285
 QY 287 QKD KALLRLMDELSSIRGVK 307
 Db 286 ---SLPDLILGELAILKGRS 303

RESULT 10

T38478
 RhoGAI/LIM domain protein - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 R:Brown, D.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, June 1997
 A:Reference number: Z1796
 A:Accession: T38478
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-969

 A:Cross-references: EMBL:Z97210; PIDN:CAB10138.1; GSPDB:GN00066; SPDB:SPAC29A4.11
 A:Experimental source: strain 972h; cosmid c29A4
 C:Genetics:
 A:Gene: SPDB:SPAC29A4.11
 A:Map position: 1

Query Match 5.8%; Score 96.5; DB 2; Length 969;
 Best Local Similarity 18.5%; Pred. No. 10;
 Matches 60; Conservative 52; Mismatches 99; Indels 113; Gaps 15;

QY 17 ASPAKTLFARYSKLTKTERKEALLTVTKANLRNMTLHYIIGHICPL---YRFSSSIV 73
 Db 321 ASPKYT-----MSLTDRAPIV-----MNGHMRSLHNATSPFRFPSPYSSDTHS 366
 QY 74 PLATHPDVMDVFTPPQKEFEIGELVKTHQRTSFHFNQFTLT---SPKESVTKNAV 130
 Db 367 PTRSPN-----VQTHK--KYSSQPSDISSFAQLLSPQVLSPPKPN 406
 QY 131 DMAYHMLEAMGIADRSVINIHIGG-----AYGNKDTATAQFH---QNI 172
 Db 407 CGCHKFRSHLSLSETSQOTLVPSLGSNGEYHLPTNDHSSTPAQSERDSIDVELREOLENL 466
 QY 173 KOLPQEIKEKMT-----LENDKTYTTEETLOVCEQEDVPFDFHFFHYANPDH 223
 Db 467 TALTKKLERLSSTSDNSKFIKRTEDKDTVRSKLEICEK-----FFSFADVTD 518

QY 224 DLNVALPRMIKTWERIGLQPKVHLSPPKSEQAIRSHADYV-DANFLERFRQ-WGTNIDF 281
DB 519 -----LKDPK-----HODLVAAANAMAMLRRESYGTETNN 548
QY 282 MIEAKOKKALLRLMDLSSIRGV 305
DB 549 LLERRNE-----LLDDYNNVOKI 566
RESULT 11
T22200
hypothetical protein F44G3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22200
R:McMurray, A.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19529
A:Accession: T22200
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-775 <WIL>
A:Cross-references: EMBL:Z83109; PIDN:CAB05518.1; GSPDB:GN00023; CESP:F44G3.3
A:Experimental source: clone F44G3
C:Genetics:
A:Gene: CESP:F44G3.3
A:Map position: 5
A:Introns: 23/1: 392/3
Query Match 5.7%; Score 95; DB 2: Length 775;
Best Local Similarity 21.3%; Pred. No. 9.8; Mismatches 135; Indels 56; Gaps 12;
Matches 64; Conservative 46
QY 28 YSKLSKTERKEALITVTKANLRTMRTLHYIIGHGIPLYRFSSSIPLATHPDVMDVFT 87
DB 476 FRKLSWIELK-----GEADDYNTIPOMR-VNGRNLTNLRFAADDIVLIANHPNTASKMLQ 528
QY 88 PFQKEFREIGELV---KTHQLRTSFHPNQFTLTSPKESVTKNVATDMAYHYRMLAMGI 144
DB 529 ELVKCSEVGLIEGTGKTVLRNRFADPSKYVFGSPSTTQDDVDEYIYLGROINQA-- 586
QY 145 ADRSVINHIHGAYGNKDTATQAFHONIKOLPOEIKERMTLEN-----DDKTYTTEE- 196
DB 587 -NNLMPEIH-----RRRAAAAFN-GIKNTDTSITDKKIRANLFDISIVLPALTYGSEAW 639
QY 197 --TLQVCEQEDVPVDFHHFYANPDHDLNVALPRMIKTWERIGLQPKVHLSPPKSEQ 254
DB 640 TTKALSERVRI-----THASLERLVLGTLTQOR---ERDLHREDIRTMS 682
QY 255 AIRSHADYVDANFLERFRQWGTNIDFWIEAKQDKALLRLMDLSSIRGVKRTGGALQ 314
DB 683 LVRD-----PLNFVKRKRKLGWAGHV-----ARRKDGRTTLMTEWRPYGWRKRPVGRPPMR 732
QY 315 W 315
DB 733 W 733
RESULT 12
T21233
hypothetical protein F22B3.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21233
R:Cottage, A.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z19394
A:Accession: T21233
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-553 <WIL>
A:Cross-references: EMBL:Z68336; PIDN:CAA92738.1; GSPDB:GN00022; CESP:F22B3.6
A:Experimental source: clone F22B3
C:Genetics:
A:Gene: CESP:F22B3.6
A:Map position: 4
Query Match 5.7%; Score 94.5; DB 2: Length 553;
Best Local Similarity 20.8%; Pred. No. 6.5;
Matches 56; Conservative 41; Mismatches 123; Indels 49; Gaps 10;
QY 60 GHGIPLYRFSSSIPLATHPDVMDVFTPFQKEFREIGELV---KTHQLRTSFHPNQFTL 116
DB 279 GRNLTSLRFADDIIVLIANHPNTASKMLQELVKCSEVGLIEGTGKTVLRNRFADPSKVY 338
QY 117 FTSKESVTKNVATDMAYHYRMLAMGIADRSVINIHIGGAYGNKDTATQAFHONIKQLP 176
DB 339 FSSPSPTTQDDVDEYIYLGROINQA---NNLMPEIH-----RRRAAAAFN-GIKNTT 389
QY 177 OEIKERMTLEN-----DDKTYTTEE---TLQVCEQEDVPVDFHHFYANPDHDLN 226
DB 390 DSITDKKIRANLFDISIVLPALTYGSEAWTTKALSERVRI-----THASLE 435
QY 227 VALPRMIKTWERIGLQPKVHLSPPKSEQAIRSHADYVDANFLERFRQWGTNIDFWIEAK 286
DB 436 RRLVGTITLTQOR---ERDLHREDIRTMSLVRD-----PLNFVKRKRKLGWAGHV-----AR 482
QY 287 QKDKALLRLMDLSSIRGVKRTGGALOW 315
DB 483 RKDGRWTTLMTEWRPYGWRKRPVGRPPMR 511
RESULT 13
A59251
myosin - Acetabularia cliftonii
C:Species: Acetabularia cliftonii
C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 08-Sep-2000
C:Accession: A59251
R:Vuqtek, O.; Menzel, D.
submitted to GenBank, March 1997
A:Description: Molecular analysis of the cDNA coding for an unconventional myosin fro
A:Reference number: A59251
A:Accession: A59251
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1145 <VUG>
A:Cross-references: GB:U94398; NID:g2051982; PIDN:AAB53062.1; PID:g2051983
C:Genetics:
A:Gene: myo2
A:Genetic code: SGC5
C:Superfamily: myosin motor domain homology
F:97-799/Domain: myosin motor domain homology <MMO>
Query Match 5.7%; Score 94.5; DB 2: Length 1145;
Best Local Similarity 20.1%; Pred. No. 19;
Matches 55; Conservative 47; Mismatches 100; Indels 71; Gaps 11;
QY 50 NTMRTLHYIIGHGIPLYRFSSSIPLATHPDVMDVFTPFQKEF-----REIGELV 100
DB 349 NAMRTIGMTGEODSVFRIVAAILHLGN-----ITPMGDEAEFSGSEAEESQACADLL 403
QY 101 K---THQLRTSFHPNQFTLTSPKESVTKNVATDMAYHYRMLAMGIADRSVINIHIGGAY 158
DB 404 KIPVQQLKT-----ALTKRNLKNTA-----GIIT---VTPLKVPAAE 436
QY 159 GNKDTATQAFHONI-KOLPOEIKERMTLENDDKTYTTEETLQVCEQEDVPVDFHHFYA 217
DB 437 ESRDALAKTIYSRLFDMLVSAIKEISFFRDTKTSATSDRTIGILD-----IYGFSEFK 490
QY 218 NPDDHADNVALPRMIKTWERIGLQPKVHLSPPKSEQAIR-----SHADYVDANFLERF 272
DB 491 NSFEJLCINLAN-----EKLOQGFNHHVLEGEQOQYIAEGISWSYVDFVDNQDCLDLL 543

QY 273 RQWGTNIDFMIEAKOKKALLRLMDELSSIRGV 305
 Db 544 EGG-----KNAKGIFFLIDEACKMPW 567

RESULT 14

T32377
 hypothetical protein K10F12.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 01-Dec-2000
 C:Accession: T32377; T22150; T16200; T22352
 R:Wohlmann, P.; Beck, C.
 submitted to the EMBL Data Library, September 1997
 A:Description: The sequence of C. elegans cosmid K10F12.
 A:Reference number: Z21157
 A:Accession: T32377
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-805 <WOH>
 A:Cross-references: EMBL:AF025462; PIDN:AAB71003.1; GSPDB:GN00021; CESP:K10F12.5
 A:Experimental source: strain Bristol N2; clone K10F12
 R:Kershaw, J.
 submitted to the EMBL Data Library, September 1996
 A:Reference number: Z19523
 A:Accession: T22150
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-120, 'F', 122-280, 'P', 282-551, 'T', 553-577, 'A', 579-805 <WIL>
 A:Cross-references: EMBL:Z79755; PIDN:CAB02112.1; GSPDB:GN00019; CESP:F43G9.7
 A:Experimental source: clone F43G9
 R:Le, T.
 submitted to the EMBL Data Library, November 1995
 A:Description: The sequence of C. elegans cosmid F28E10.
 A:Reference number: Z18476
 A:Accession: T16200
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 253-551, 'T', 553-805 <LET>
 A:Cross-references: EMBL:U040801; NID:g1065958; PID:g1065959; PIDN:AAA81500.1; CESP:F28E10
 R:Mathews, L.
 submitted to the EMBL Data Library, August 1996
 A:Reference number: Z19552
 A:Accession: T22352
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 253-551, 'T', 553-697, 'L', 699-805 <WT2>
 A:Cross-references: EMBL:Z79599; PIDN:CAB01870.1; GSPDB:GN00028; CESP:F47C8.1
 A:Experimental source: clone F47C8
 C:Genetics:
 A:Gene: CESP:K10F12.5; CESP:F43G9.7; CESP:F28E10.3; CESP:F47C8.1
 A:Map position: 3

Query Match 5.6%; Score 93.5; DB 2; Length 805;
 Best Local Similarity 20.8%; Pred. No. 14;
 Matches 56; Conservative 41; Mismatches 123; Indels 49; Gaps 10;
 QY 60 GHGIPLYRFESSIVPLATHPDVMDVFTPOKEFREIGELV---KTHQLRTSFHPNOFTL 116
 Db 531 GRNLTLRFADDIVLIANHPNIASMKLOELVOKSEVGLINTGKTKVLNRNRFADPSKVY 590
 QY 117 FTSPKESVTKNAVTDMAHYHRLMEAMGIADRSVINIHIGGAYGNKDTATAOFTONIKQLP 176
 Db 591 FGSPSPPTQLDDVDYIVLGRQINAQ---NNLMPEIH-----RRRAAWAFN-GIKNTT 641
 QY 177 QFIKERTLEN-----DKTYTTEE---TLQVCEQEDVFFDFHHFYANPDHDLN 226
 Db 642 DSITDKKIRANLFDISIVLPALTGSEAWTFTKALSERVRI-----THASLE 687
 QY 227 VALPRMKTWERIGLOPKVHLSFKSQAIRSHADYVDANFLERFROWGTNIDEMIEAK 286
 Db 688 RRLVGIITLQOR---ERDLHREDIRTSLYRD-----PLNFVKRKLGNAGHV-----AR 734

QY 287 QKDKALLRLMDELSSIRGVKRIIGGALQW 315
 Db 735 RKDGRWTTLTMTWPRPYGWKRPVGRPPMRW 763

RESULT 15

F71565
 probable isoleucine--tRNA ligase (EC 6.1.1.5) - Chlamydia trachomatis (serotype D, st
 C:Species: Chlamydia trachomatis
 C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 24-Nov-1999
 C:Accession: F71565
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
 Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
 A:Reference number: A71570; MUID:99000809
 A:Accession: F71565
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1036 <ARN>
 A:Cross-references: GB:AE001277; GB:AE001273; NID:g3328408; PIDN:AAC67609.1; PID:g332
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: files
 C:Superfamily: isoleucine--tRNA ligase
 C:Keywords: ligase

Query Match 5.6%; Score 93.5; DB 2; Length 1036;
 Best Local Similarity 19.7%; Pred. No. 20;
 Matches 57; Conservative 42; Mismatches 104; Indels 87; Gaps 13;
 QY 71 SIIVPLATHPDVMDV-----DFVT-----PFQKFRFREGELVKTHQLRTSFHP 111
 Db 229 SNMALAVHPELTYVRKDKESGDEYILGQESLPWFDPRESYEWIGOLSGKSLVQGSYEP 288
 QY 112 NQFTLTSPKESVTKNAVTDMAHYHRLMEAMGIADRSVIN-----IHIGGAYGNKDTA 164
 Db 289 -LFPYEQDKKE-----LEAFRLPADFTIESESEGTGIVHMAPAFGEADFF 331
 QY 165 TAQFH-----QNIKQLPOEIKERTMLENDKTY---TTEFTL 198
 Db 332 ACEHNVPVPCVDNOCGYTAEVKDFVGEYIKSADKGIARR--LKNENKLFYQGTVRHRY 389
 QY 199 QVCEQEDVFFDFHHFYANPDHDLNVALPRMKTWERIGLOPKVHLSFKSQAIRS 258
 Db 390 PFCWRTDPSLIYKAVNSW-----FVAVERKVKSKMLKANESIHTWPE-HIKQGRFGKNLEG 443
 QY 259 HADYVDANELLERFROWGTNIDFMIEAKOKKALLRL-----MDELSIRGVK 306
 Db 444 ARDWA-----ISRNRYWGTPIPIW---RSDDGELLVIGSIQLEALSQK 485

Search completed: January 15, 2002, 13:52:32
 Job time: 401 sec

us-09-724-296-37.rpr

Tue Jan 15 14:12:14 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:07:17 ; Search time 45.07 seconds
(without alignments)
257.882 Million cell updates/sec

Title: US-09-724-296-37

Perfect score: 1662

Sequence: 1 MIFRFGVSNMSLWDASPA.....ELSSINGVKRIGGALQWKS (317)

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1650.5	99.3	320	YWJD_BACSU	P45864 bacillus su
2	103	6.2	964	CTA2_BACCI	P70873 bacillus ci
3	93.5	5.6	1036	1 SYI_CHLTR	O84022 chlamydia t
4	92.5	5.6	501	1 C4D2_DROME	Q27589 drosophila
5	91.5	5.5	522	1 SNX1_HUMAN	Q13596 homo sapien
6	91.5	5.5	1409	1 COPI_DROME	P04146 drosophila
7	91.5	5.5	2334	1 WAPA_BACSU	Q07833 bacillus su
8	91	5.5	1284	1 ATI_COMPLEX	P16602 cowpox viru
9	90.5	5.4	233	1 RLL1_THEMA	P29393 thermotoga
10	90	5.4	572	1 GAG_IPHA	P04023 hamster int
11	90	5.4	1941	1 MYH2_HUMAN	Q9UKX2 homo sapien
12	90	5.4	2104	1 MYS3_SCHPO	O14157 schizosacch
13	89.5	5.4	429	1 APA2_MACFA	P33621 macaca fasc
14	89.5	5.4	5255	1 BACA_BACLI	O68006 b bacitraci
15	89	5.4	550	1 SYD_MYCGE	P47282 mycoplasma
16	89	5.4	1033	1 TIRI_ECOLI	P10486 escherichia
17	88.5	5.3	582	1 HEMQ_OPSTA	P43090 opsanus tau
18	87.5	5.3	856	1 VPH1_NEUCR	Q01290 neurospora
19	86.5	5.2	533	1 KCED_RAT	P15791 rattus norv
20	86.5	5.2	802	1 PUR2_YEAST	P07244 s bifunctio
21	86.5	5.2	1259	1 LIN1_HUMAN	P08547 homo sapien
22	86	5.2	427	1 RHO_THEMA	P38527 thermotoga
23	86	5.2	583	1 FOJO_DROME	P54360 drosophila
24	86	5.2	1083	1 UL52_VZVD	P09270 varicella-z
25	86	5.2	1131	1 MYPE_CHICK	P16419 gallus gall
26	85	5.1	326	1 HEM2_PASMU	P57874 pasteurella
27	85	5.1	601	1 HEM1_MOUSE	O62431 mus musculu
28	84.5	5.1	382	1 MATB_NEUCR	P36981 neurospora
29	84.5	5.1	399	1 RPSD_THEMA	P77994 thermotoga
30	84.5	5.1	508	1 CPVL_RAT	P22443 rattus norv
31	84.5	5.1	528	1 MASY_EMENI	P28344 emeritella
32	84.5	5.1	635	1 SYT_CHLPN	Q9Z7A0 chlamydia p
33	84.5	5.1	1085	1 CUT7_SCHPO	P24339 schizosacch

How did they know so definitely?

NO PA

34	84.5	5.1	1679	1 YIO9_YEAST	P40457 saccharomyc
35	84.5	5.1	1739	1 CHD2_HUMAN	O14647 homo sapien
36	84	5.1	276	1 COX2_HUMAN	O9Y6N1 homo sapien
37	84	5.1	649	1 METX_YEAST	Q04533 saccharomyc
38	84	5.1	698	1 SST2_YEAST	P11972 saccharomyc
39	84	5.1	724	1 ATIL_VACCV	P24759 vaccinia vi
40	84	5.1	843	1 YMS1_YEAST	Q05050 saccharomyc
41	83.5	5.0	436	1 EP45_XENLA	Q00387 xenopus lae
42	83.5	5.0	829	1 GUTR_BACSU	P39143 bacillus su
43	83.5	5.0	1251	1 RBP2_PLAVB	Q00799 plasmodium
44	83	5.0	343	1 METE_THEAC	P57704 thermoplasm
45	83	5.0	944	1 NUF1_YEAST	P32380 saccharomyc

ALIGNMENTS

RESULT	1
YWJD_BACSU	
ID	YWJD_BACSU
AC	P45864:
DT	01-NOV-1995 (Rel. 32, Created)
DT	01-NOV-1995 (Rel. 32, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	HYPOTHETICAL 36.9 KDA PROTEIN IN ACDA 5' REGION.
GN	YWJD.
OS	Bacillus subtilis.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;
OC	Bacillus/Staphylococcus group; Bacillus.
OX	NCBI_TaxID=1423;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=168;
RA	Glaser P., de la Fuente V., Danchin A.;
RL	Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	
DR	EMBL; Z49782; CAA89865.1; -
DR	EMBL; Z99123; CAB15748.1; -
DR	Subtilist; BG11309; YWJD.
KW	Hypothetical protein; Complete proteome.
SQ	SEQUENCE 320 AA; 36896 MW; 2FD35A710648E002 CRC64;

Query Match	99.3%	Score	1650.5;	DB 1:	Length	320;	
Best Local Similarity	99.1%	Pred. No.	1.1e-127;				
Matches	317;	Conservative	0;	Mismatches	0;	Gaps	1;
Qy	1	MIFRFGVSNMSLWDASPAKTLTFARYSKLSKTERKEALLTVTKANLRNTRTLHYIG	60				
Db	1	MIFRFGVSNMSLWDASPAKTLTFARYSKLSKTERKEALLTVTKANLRNTRTLHYIG	60				
Qy	61	HGIPLYRFSSSVPLATHPDVMDVFTPPQKEFREGELVKTHOLRTPSPHNOFTLTSP	120				
Db	61	HGIPLYRFSSSVPLATHPDVMDVFTPPQKEFREGELVKTHOLRTPSPHNOFTLTSP	120				
Qy	121	KESVTKNAVTDMAHYHRLMEAMGADRSVINIHIGAYGNKDTATATQFHNKQLPQETK	180				
Db	121	KESVTKNAVTDMAHYHRLMEAMGADRSVINIHIGAYGNKDTATATQFHNKQLPQETK	180				
Qy	181	ERMTLENDKTYTTEETLQVCEQEDVPFVDFHHFYANPDHADLNVALPRMKTWERTG	240				
Db	181	ERMTLENDKTYTTEETLQVCEQEDVPFVDFHHFYANPDHADLNVALPRMKTWERTG	240				
Qy	241	LQPKVHLSSPKSEQAIRSHADYVDANF--LLERFRQWGTNIDFMIEAKQKDKALLRLMD	297				

```

Db 241 LQPKVHLLSPKSEQATIRSHADYVDANFLLPFLERFROWGNTIDFMIEAKOKDKALLRLMD 300
QY 298 ELSSIRGVKRIIGGALOWKS 317
Db 301 ELSSIRGVKRIIGGALOWKS 320

RESULT 2
CTA2_BACCI STANDARD; PRT; 964 AA.
ID P70873;
AC 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE CYCLOISOMALTOOLIGOSACCHARIDE GLUCANOTRANSFERASE PRECURSOR (EC 2.4.1.-)
DE (CITASE).
OS Bacillus circulans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1397;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=U-155;
RA Oguma T., Kurokawa T., Tobe K., Kitao S., Kobayashi M.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC FUNCTION: PRODUCES CYCLOISOMALTOOLIGOSACCHARIDE FROM DEXTRAN.
CC -!- SIMILARITY: BELONGS TO FAMILY 66 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D88360; BAA13595.1;
KW Transferase; Glycosyltransferase; Signal.
FT SIGNAL 30 POTENTIAL.
FT CHAIN 31 964 CYCLOISOMALTOOLIGOSACCHARIDE
FT GLUCANOTRANSFERASE.
FT SEQUENCE 964 AA; 107208 MW; 8849CDC3E2DE9A68 CRC64;

Query Match 6.2%; Score 103; DB 1; Length 964;
Best Local Similarity 22.1%; Pred. No. 1.2;
Matches 55; Conservative 44; Mismatches 84; Indels 66; Gaps 14;

QY 95 EIGELVKTHTLRTSPHNPQFTLTSPKSVTKNAVDMAYHYRMLEAMG--IADRSVIN- 151
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
Db 561 ELGE--DSQMLAHEYYPNR-----SKSMRSTLKSAMKD---HYNFTAYENLLFDADVIDN 611
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
QY 152 -----IHIGGAYGNKDTATAQFONIKOLPO--ELKERMITLENDKTY-----TTEET- 197
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
Db 612 DAGKQFINTAGVNTSPDGAANTVWHSKRTPEYNTLHLINLVNNDQWNSNQTAQTN 671
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
QY 198 ----LQVCEQEDVPFVDFHFHYANPDDHADNLVALP-----RMKWTWE 237
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
Db 672 LATKVIYIGAEETITGV-----YAAPDHNOGATQSLPFTTGTDSGYSFTSVPSLEYWS 726
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
QY 238 RIGLQPKVHLSSP-----KSEQAIRS-----HADYVDANFLLERFROWGNTIDFMIEA 285
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
Db 727 MIYN--KRSTAAAPVDNMYEAEATAIKSNVSVNTNHAGYTGSGF--VDQFATVNDGVSVFIVHA 783
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
QY 286 KOKDKALLR 294
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
Db 784 SSKDDYVLR 792
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|

RESULT 3
SYL_CHLTR
ID SYL_CHLTR STANDARD; PRT; 1036 AA.

```

```

084022;
20-AUG-2001 (Rel. 40, Created)
20-AUG-2001 (Rel. 40, Last sequence update)
20-AUG-2001 (Rel. 40, Last annotation update)
DE ISOLEUCYL-TRNA SYNTHETASE (EC 6.1.1.5) (ISOLEUCINE--TRNA LIGASE)
DE (ILERS).
GN LES OR CT019.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -!- CATALYTIC ACTIVITY: ATP + L-ISOLEUCINE + TRNA(ILE) = AMP +
CC PYROPHOSPHATE + L-ISOLEUCYL-TRNA(ILE).
CC -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001277; AAC67609.1;
DR HSSP; P56690; 11LE.
DR InterPro; IPR002300; trna-synt_la.
DR InterPro; IPR001412; trna-synt_l.
DR InterPro; IPR002301; trna-synt_ile.
DR Pfam; PF00133; trna-synt_1; 1.
DR PRINTS; PR00984; TRNASYNTHILE.
DR PROSITE; PS00178; AA-TRNA-LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Metal-binding; Zinc; Complete proteome.
FT SITE 46 56 "HIGH" REGION.
FT SITE 589 593 "KMSKS" REGION.
FT BINDING 592 592 ATP (BY SIMILARITY).
FT SEQUENCE 1036 AA; 118985 MW; DC4420FD3734EFD CRC64;

Query Match 5.6%; Score 93.5; DB 1; Length 1036;
Best Local Similarity 19.7%; Pred. No. 8;
Matches 57; Conservative 42; Mismatches 104; Indels 87; Gaps 13;

QY 71 SIPLVATHPDVMMW-----DFVT-----PFOKEFREIGELVKTHTLRTSFP 111
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
Db 229 SNMALAVHPELTYVRIKDKESGDEYILGOESLPRWFPDRESYEWIGSLGSLVGQSYEP 288
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
QY 112 NOFTLTSPKESVTKNAVDMAYHYRMLEAMGADRSVIN-----IHIGGAYGNKDTA 164
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
Db 289 -LFPYFQPKKE-----LEAFRIPLADFIESEGTGIVHMAPAFGEADFF 331
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
QY 165 TAOEH-----ONIKOLPOELIKERMITLENDKTY---TTEETL 198
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
Db 332 ACQEHNVPLVCPVDNNOGCTAEVKDFVGEYIKLSADKGIARR--LKNENLFYQCTVRRHY 389
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
QY 199 QVCEQEDVPFVDFHFHYANPDDHADNLVALPRLIKTWERIGLQPKVHLSPKSEQAIRS 258
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
Db 390 PFCWRTDSPLIYKAVNSW-----FVAVEKVKSKMLKANESIHWTPE-HIKOGRGKWLKG 443
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|
QY 259 HADYVDANFLLERFROWGNTIDFMIEAKQDKALLRL--MDELSIRGVK 306
|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:| :| :|:|:|

```

Db 444 ARDWA-----ISRNRWGTPIPIW---RSDGELLVIGSIQLEALSQK 485

RESULT 4

C4D2_DROME STANDARD; PRT; 501 AA.

AC Q27589; O27588; O46053; O18651; O18674; Q9W514;

DT 15-DEC-1998 (Rel. 37, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE CYTOCHROME P450 4D2 (EC 1.14.1.1) (CYP1B2).

GN CYP4D2 OR EG:152A3.4 OR CG3466.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OREGON-R;

RX MEDLINE=94296569; PubMed=8024706;

RA Frolow M.V., Alatorsev V.E.;

RT "Cluster of cytochrome P450 genes on the X chromosome of Drosophila melanogaster.";

RL DNA Cell Biol. 13:663-668(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=OREGON-R;

RX MEDLINE=20196011; PubMed=10731137;

RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,

RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,

RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,

RA Minana B., Kafatos C., Louis C., Siden-Kiamos I., Bolshakov S.,

RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablo B.,

RA Modolell J., Peter A., Schoettler P., Werner M., Mourikioti F.,

RA Beinert N., Dowe G., Schaefer U., Jaekle H., Bucheton A.,

RA Callister D.M., Campbell L.A., Darlamsou A., Henderson N.S.,

RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,

RA Glover D.M.;

RT "From sequence to chromosome: the tip of the X chromosome of D. melanogaster.";

RL Science 287:2220-2222(2000).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA April J.F., Agbayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,

RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Besu P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Dawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Fieser K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karp G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulyov G., Hishina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.R., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

RN [4]

RP SEQUENCE OF 4-501 FROM N.A.

RC STRAIN-VARIOUS STRAINS;

RA Phillips K.S., Begun D.J., Aquadro C.F.;

RT "Evidence for non-neutral evolution around the cytochrome p450 gene cluster on the Drosophila melanogaster X chromosome.";

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: INVOLVED IN THE METABOLISM OF INSECT HORMONES AND IN THE BREAKDOWN OF SYNTHETIC INSECTICIDES (BY SIMILARITY).

CC -!- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) = ROH + OXIDIZED FLAVOPROTEIN + H(2)O.

CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM (POTENTIAL).

CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X75955; CAA53568.1; -

CC EMBL; Z23005; CAA80549.1; -

CC EMBL; AL009194; CAA15698.1; -

CC EMBL; AE003423; AAF45741.1; -

CC EMBL; AF017006; AAB71169.1; -

CC EMBL; AF017007; AAB71170.1; -

CC EMBL; AF017008; AAB71171.1; -

CC EMBL; AF017009; AAB71172.1; -

CC EMBL; AF017010; AAB71173.1; -

CC EMBL; AF017011; AAB71174.1; -

CC EMBL; AF017012; AAB71175.1; -

CC EMBL; AF017013; AAB71176.1; -

CC EMBL; AF017014; AAB71177.1; -

CC EMBL; AF017015; AAB71178.1; -

CC EMBL; AF017016; AAB71179.1; -

CC EMBL; AF017017; AAB71180.1; -

CC EMBL; AF017018; AAB71181.1; -

CC FlyBase; FBgn0011576; Cyp4d2.

CC InterPro; IPR001128; Cyt_P450.

CC Pfam; PF00067; p450; 1.

CC PRINTS; PR00385; P450.

CC PRINTS; PR00464; EP450II.

CC PROSITE; PS00086; CYTOCHROME_P450; 1.

CC Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;

CC Endoplasmic reticulum; Polymorphism.

FT BINDING 449 449 HEME (BY SIMILARITY).

FT BINDING 163 163 K -> M (IN STRAINS CAM-44, CAM-48 AND BERKELEY).

FT CONFLICT 2 25 MISSING (IN REF. 2).

FT CONFLICT 30 30 I -> A (IN REF. 1; CAA80549).

FT CONFLICT 160 160 A -> R (IN REF. 1).

FT CONFLICT 487 501 LRSANGVHLGKPR -> CGRPTAFILA (IN REF. 1).

FT SQ SEQUENCE 501 AA; 57914 MW; 8F8D98DBF39FDIA2 CRC64;

Query Match 5.6%; Score 92.5; DB 1; Length 501;

Best Local Similarity 19.8%; Pred. No. 3.6;

Matches 55; Conservative 43; Mismatches 91; Indels 89; Gaps 12;

60 GHGI-----PLYRESSIVPLATHPDVMDVFPFOKEFREIGELVKTOLRT-SFHPNQ 113
113
27 GNLGLPGRPLPFFLGNLLMYRGDPEQIMDFVKKORKYGRGLYRVWLLHOLAVFSTDPD 86
86
114 FTLFTSPKESVTKNAVDMAYHYRMLEAMGIAIDRSVNIHIGGA-----YGNKDTA 164
164
87 IEFVLSQOHTTKNNL-----YKLL-----NCWLGDLMLSTGRKWHGRKKII 129
129
165 TAOFHONIKOLPOEIKERMTLENDKTYTTEELVQCEQEDVFEVDFHFFHYANPDH 224
224
130 TPTTFHKI-----LEQFVEIFDQGSAMV-----EQLORADGKTP 165
165
225 LNV-----ALPRMIKT-WERTIGLOPKVHLSSPKSEQAIRSHADYVDA-----NFLLE 270
270
166 INIEPVLITLALDIAETAMGCTKNAOKNPNL-----PVQAVNDVTNLIK 212
212
271 RFRWGTNIDPMI-----EAKQDKALLRLMDELSSI 302
302
213 RFIHAWQRVDWIFRLTOPTTEAKRQDKAIVKMHDFTEI 250
250
RESULT 5
SNX1 HUMAN STANDARD; PRT: 522 AA.
ID SNX1 HUMAN STANDARD; PRT: 522 AA.
AC Q13596; 060750; 060751;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SORTING NEXIN 1.
OS Homo sapiens (Human).
GN Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96208617; PubMed=8638121;
RA Kurten R.C., Cadena D.L., Gill G.N.;
RT "Enhanced degradation of EGF receptors by a sorting nexin, SNX1.";
RL Science 272:1008-1010(1996).
RN [2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=99038232; PubMed=9819414;
RA Haft C.R., de la Luz Sierra M., Barr V.A., Haft D.H., Taylor S.I.;
RT "Identification of a family of sorting nexin molecules and
characterization of their association with receptors.";
RL Mol. Cell. Biol. 18:7278-7287(1998).
CC -!- FUNCTION: MAY BE INVOLVED IN SEVERAL STAGES OF INTRACELLULAR
TRAFFICKING. PLAYS A ROLE IN TARGETING LIGAND-ACTIVATED EGFR TO
THE LYSSOMES FOR DEGRADATION AFTER ENDOCYTOSIS FROM THE CELL
SURFACE AND RELEASE FROM THE GOLGI.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 1A: ARE
PRODUCED BY ALTERNATIVE SPLICING.
CC -!- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PHOX HOMOMOLOGY (PX) DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: U53225; AAA98672.1;
CC EMBL: AF065483; AAC17182.1;
CC EMBL: AF065484; AAC17183.1;
CC MIM: 601272;
CC InterPro: IPR001683; PX.
CC Pfam: PF00787; PX; 1.
CC SMART: SM00312; PX; 1.
CC Transport; Protein transport; Golgi stack; Alternative splicing.
KW

FT DOMAIN 145 265 PX.
FT VARSPLIC 91 155 MISSING (IN ISOFORM 1A).
FT CONFLICT 117 117 S -> P (IN REF. 1).
FT CONFLICT 211 211 S -> P (IN REF. 1).
SQ SEQUENCE 522 AA; 59049 MW; FC8BB93977A5AE70 CRC64;

Query Match 5.5%; Score 91.5; DB 1; Length 522;
Best Local Similarity 19.3%; Pred. No. 4.6;
Matches 64; Conservative 54; Mismatches 106; Indels 107; Gaps 16;

QY 28 YSKLKTER-----KEALLTVTKANL-RNTMRTLHYLIIGHIGIPLYRESSIVPLA 76
Db 194 YEKLEKHSQNGFIWPPSPKSLIGMTKVKGKEDSSSAEFLKRAALERYLQIRV--- 250
QY 77 THPDWMDVTPFOKEFREIGELVKTOLRTSFHFNQFTLTFTSPKESVTKNAV-----TDM 132
Db 251 NIPTMLQD---FDVREFLEKEELPRAVGTOTLSGAGLLKMKENKATDAVSKMTIKMNESDI 307
QY 133 AYHYRMLEAMGIADR-----SVINIHIGGAYGNKDTA--TAOFHONIKOLPOEIKERMTL 185
Db 308 WFEKLOEVECEQRKLKLVAVVETLVN---HRKELALNTAQFAKSLAMLG----- 355
QY 186 ENDDKTYTTEETLOVCEQEDVFEVDFHFFYANPD-----DH-----ADNLVALPRM 232
Db 356 SSEDNTALSRLASQLAEVEEK---IEQLHQEQANNDFLLAEILLSDTIRLLAIVRAAFDOR 413
QY 233 IKTWIRIGLOPKVHLSSPKSEQAIRSHADYVDANFLERFRQ-----WGTNIDPMIEAK 286
Db 414 MKTQWR-----WQDAQATLOKKREAREALLWANKPDKLQQAQ 450
QY 287 QDKKALLRLMDELSSIRGVKRGIGGALQWKS 317
Db 451 -----DEI-----LEWES 458

RESULT 6
COPI1 DROME STANDARD; PRT: 1409 AA.
ID COPI1 DROME STANDARD; PRT: 1409 AA.
AC P04146; Q24585; Q24586; Q03728; Q24280; Q24555; Q24587;
DT 01-NOV-1986 (Rel. 03, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COPIA PROTEIN [CONTAINS: COPIA VLP PROTEIN; COPIA PROTEASE
DE (EC 3.4.23.-)].
DN COPIA.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RX MEDLINE=85267679; PubMed=2410772;
RA Mount S.M., Rubin G.M.;
RT "Complete nucleotide sequence of the Drosophila transposable element
RT copia: homology between copia and retroviral proteins.";
RL Mol. Cell. Biol. 5:1630-1638(1985).
RN [2]
RP SEQUENCE FROM N.A., SEQUENCE OF 2-10, AND ALTERNATIVE SPLICING.
RX MEDLINE=85240569; PubMed=2409449;
RA Emori Y., Shiba T., Kanaya S., Inouye S., Yuki S., Saigo K.;
RT "The nucleotide sequences of copia and copia-related RNA in Drosophila
RT virus-like particles";
RL Nature 315:773-776(1985).
RN [3]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RX MEDLINE=89183629; PubMed=2538806;
RA Miller K., Rosenbaum J., Zbrzezna V., Pogo A.O.;
RT "The nucleotide sequence of Drosophila melanogaster copia-specific
RT 2.1-kb mRNA.";
RL Nucleic Acids Res. 17:2134-2134(1989).
RN [4]


```

RESULT 12
MYS3_SCHPO STANDARD; PRT: 2104 AA.
AC OL157: 042730;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN TYPE II HEAVY CHAIN 2.
GN MYO3 OR MYP2 OR SPAC4A8.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9811930; PubMed=9459302;
RA Motegi F., Nakano K., Kitayama C., Yamamoto M., Mabuchi I.;
RT "Identification of Myo3, a second type-II myosin heavy chain in the
RT fission yeast Schizosaccharomyces pombe.";
RL FEBS Lett. 420:161-166(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98062346; PubMed=9398685;
RA Bezanilla M., Forsburg S.L., Pollard T.D.;
RT "Identification of a second myosin-II in Schizosaccharomyces pombe:
RT Myp2p is conditionally required for cytokinesis.";
RL Mol. Biol. Cell 8:2693-2705(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: STABILIZES THE F-ACTIN CABLES FORMING THE F-ACTIN RING
CC THAT SURROUNDS THE NUCLEUS DURING INTERPHASE. MAY WORK IN
CC CONJUNCTION WITH MYO2.
CC -!- SUBUNIT: BINDS TO CDC4 AND RLC1.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AB007633; BAA24579.1;
CC EMBL: AF029788; AAC04615.1;
CC EMBL: Z98762; CAB11475.1;
CC HSSP: P08799; 1MND.
CC InterPro: IPR001609; myosin_head.
CC Pfam: PF00063; myosin_head; 1.
CC PRINTS: PR00193; MYOSINHEAVY.
CC ProDom: PD000355; myosin_head; 1.
CC SMART: SM00242; MYSC; 1.
CC Myosin: Coiled coil; ATP-binding; Actin-binding; Alkylation.
CC DOMAIN 1 828 MYOSIN HEAD-LIKE.
CC FT FT DOMAIN 829 2104 COILED COIL (POTENTIAL).
CC FT FT DOMAIN 646 660 ACTIN-BINDING (POTENTIAL).
CC FT NP_BIND 182 189 ATP (POTENTIAL).
CC FT MOD_RES 686 686 ALKYLATION (BY SIMILARITY).
CC FT CONFLICT 1193 1193 D -> G (IN REF. 2).
CC FT CONFLICT 1304 1304 E -> G (IN REF. 2).
CC FT CONFLICT 1344 1344 E -> K (IN REF. 2).
CC FT CONFLICT 1420 1420 G -> D (IN REF. 2).
CC SEQUENCE 2104 AA; 242570 MW; 3A0548594028D258 CRC64;

```

Query Match 5.48; Score 90; DB 1; Length 2104;
 Best Local Similarity 19.84; Pred. No. 40;
 Matches 66; Conservative 58; Mismatches 130; Indels 80; Gaps 12;

```

QY 22 TLTFAVYSKLSKTERKEALLT-----VTKANLRMTMTLHVIIIGHIPLYRFSSSIPVL 75
Db 863 TITKERVQLTNDLQEQALAEKDLIVERANSR--VEVVHERL-----SSLENO 910
QY 76 ATHPDVMDVFVTPFOKEFREIGELVTKHQLTFSFHPNQFTLTSPKESVTKNNAVTDMAVH 135
Db 911 VTIADEKYEFL--YAEKOSIEEDLANKQTEISYLSDSLSTLEKLSIKKDKQT--ISSK 966
QY 136 YRMLEAMGTADRSVINI-----HIGGAYGNKDTATQAFHONIKOLPOET----- 179
Db 967 YKELE-----KDYLNIMADYQHSOHLNLEKAIINEKNINIRELNEKLMRLDDELLKKQ 1020
QY 180 -----KERMTLENDKDTYTT-----BETLQVCEQEDVPFV 209
Db 1021 RSYDTKVOQELREENASLKDQCTYESOLASLSVSKYSESELNKKAEELVFOKEITEYR 1080
QY 210 PDFHHFYANPDHDADNLVALPRMIKWERICLOPKVHLSPPKSEQAIRS--HADVVDANFL 268
Db 1081 DQLHKAFQNPETHNINDVKSGLNSDENIYSTSTLTSLKDVQVQLKSJHTK--EANQL 1138
QY 269 LERFRQWGTNIDFMLEAKOKDKALLRLMLDELSI 302
Db 1139 SERIKEISEMLEQSIATEEK---LRRKNSLDCI 1169

RESULT 13
AP04_MACFA STANDARD; PRT: 429 AA.
AC P33621;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE APOLIPOPROTEIN A-IV PRECURSOR (APO-AIV).
GN APOA4.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leukocyte;
RX MEDLINE=93192330; PubMed=8448212;
RA Osada J., Poci M., Nicolosi R.J., Schaefer E.J., Ordovas J.M.;
RT "Nucleotide sequences of the Macaca fascicularis apolipoprotein C-III
RT and A-IV genes.";
RL Biochim. Biophys. Acta 1172:335-339(1993).
CC -!- FUNCTION: MAY HAVE A ROLE IN CHYLOMICRONS AND VLDL SECRETION AND
CC CATABOLISM. REQUIRED FOR EFFICIENT ACTIVATION OF LIPOPROTEIN
CC LIPASE BY APOC-II; POTENT ACTIVATOR OF LCAT. APOA-IV IS A MAJOR
CC COMPONENT OF HDL AND CHYLOMICRONS.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- TISSUE SPECIFICITY: SECRETED IN PLASMA.
CC -!- DOMAIN: NINE OF THE THIRTEEN 22-AMINO ACID TANDEM REPEATS (EACH
CC 22-MER IS ACTUALLY A TANDEM ARRAY OF TWO, A AND B, RELATED 11-
CC MERS) OCCURRING IN THIS SEQUENCE ARE PREDICTED TO BE HIGHLY ALPHA-
CC HELICAL, AND MANY OF THESE HELICES ARE AMPHIPATHIC. THEY MAY
CC THEREFORE SERVE AS LIPID-BINDING DOMAINS WITH LECITHIN:CHOLESTEROL
CC ACYLTRANSFERASE (LCAT) ACTIVATING ABILITIES.
CC -!- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X68361; CAA48421.1;
CC PIR: S29565; S29565.
CC PIR: S30195; S30195.
CC HSSP: P02649; INFO.

```

```

DR InterPro: IPR000074; Apolipoprotein.
KW Pfam: PF01442; Apolipoprotein; 1.
KW Plasma; Lipid transport; HDL; Chylomicron; Repeat; Signal.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 429 APOLIPOPROTEIN A-IV.
FT DOMAIN 33 330 13 X 22 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 33 330 1.
FT REPEAT 33 330 2.
FT REPEAT 60 81 3.
FT REPEAT 82 103 4.
FT REPEAT 115 136 5.
FT REPEAT 137 158 6.
FT REPEAT 159 180 7.
FT REPEAT 181 202 8.
FT REPEAT 203 224 9.
FT REPEAT 225 246 10.
FT REPEAT 247 268 11.
FT REPEAT 269 286 12.
FT REPEAT 287 308 13.
FT REPEAT 309 330 14.
FT DOMAIN 372 420 GLU/GLN-RICH.
FT SEQUENCE 429 AA: 49876 MW: 3D458F551D0DB60C CRC64:
SQ

Query Match 5.4%; Score 89.5; DB 1; Length 429;
Best Local Similarity 20.4%; Pred No. 5.2;
Matches 68; Conservative 59; Mismatches 142; Indels 65; Gaps 13;
QY 8 VSNMGLNDASPAKLTFTFARYSKLSKTERKEALLTVTKANLRNTMRTLHYIIGHGIPLYR 67
Db 15 VTGAREVSGADQVATVMDYFSQLS-SNAKEAVEHLQKSELQTQQLNALFDQKLGENVYIA 73
QY 68 --FSSSIVPLAT--HPDVMD--FVTPQKEFRE-----GELVKTHQL 105
Db 74 GDQKLVFATLHERLAKOSEKLEIRKEEVRALLPHANEVSQKIGENVRELQQ 133
QY 106 RTSFHPNQFTLTPSKESYTKNAVTDMAHY-RMLEAMGIADRSVINIHIGAVGNKDTA 164
Db 134 RLEPVTQDLRTQVNTQTEQLRRLQTPYAQRMERVLRENADSLQSLRPH-----ADQL 186
QY 165 TQAFHONIKQLPQELKERTMLENDKTYTTBETLOVCEQEDVPFVDFPHHYANPDHAD 224
Db 187 KAKIQNV---EELKERLTPYADEFKYKIDQTVTEELRSLAPYA-----QDAQEK 233
QY 225 LNVALP---RMKWTWIRIGLPKQVHLSPKSEQAIRSHADYVDANF-----LLERFQ 274
Db 234 LNHLQGLAFQMKNAEB--LKARISASAEELRQLAPLAEDMRGNLGNTEGLQKSLAE 291
QY 275 WGTNIDEMIE-----AKOKKALLRLMDEL 299
Db 292 LGGHLDRHVEBFRLRVEPYGENFNKALVQOMEQL 325

RESULT 14
BACA_BACLI STANDARD; PRT; 5255 AA.
AC O68006;
AT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE BACITRACIN SYNTHETASE 1 (BAC) [INCLUDES: ATP-DEPENDENT ISOLEUCINE
DE ADENYLASE (ILEA) (ISOLEUCINE ACTIVASE); ATP-DEPENDENT CYSTEINE
DE ADENYLASE (CYSA) (CYSTEINE ACTIVASE); ATP-DEPENDENT LEUCINE ADENYLASE
DE (LEUA) (LEUCINE ACTIVASE); ATP-DEPENDENT GLUTAMATE ADENYLASE (GLUA)
DE (GLUTAMATE ACTIVASE); ATP-DEPENDENT ISOLEUCINE ADENYLASE (ILEA)
DE (ISOLEUCINE ACTIVASE); GLUTAMATE RACEMASE (EC 5.1.1.3)].
GN BACA.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group.
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_taxid=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10716;

```

MEDLINE=98089193; PubMed=9427658;
 Konz D., Klens A., Schoerndorfer K., Marahiel M.A.:
 "The bacitracin biosynthesis operon of Bacillus licheniformis ATCC
 10716: molecular characterization of three multi-modular peptide
 synthetases.";
 Chem. Biol. 4:927-937(1997).
 CC -!- FUNCTION: ACTIVATES FIVE AMINO ACIDS, INCORPORATES TWO D-AMINO
 ACIDS, RELEASES AND CYCLIZES THE MATURE BACITRACIN.
 CC -!- CATALYTIC ACTIVITY: L-GLUTAMATE = D-GLUTAMATE.
 CC -!- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTHETHEINES
 (POTENTIAL).
 CC -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
 ANTIBIOTIC BACITRACIN.
 CC -!- SUBUNIT: LARGE MULTIMERIC COMPLEX OF BAI, BA2 AND BA3.
 CC -!- DOMAIN: CONSISTS OF FIVE MODULES AND ONE EPIMERIZATION DOMAIN IN
 THE FOURTH MODULE. EACH MODULE INCORPORATES ONE AMINO ACID INTO
 THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS
 RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOATION, CONDENSATION
 (NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND
 N METHYLATION (OPTIONAL).
 CC -!- MISCELLANEOUS: BACITRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC
 DODECAPEPTIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST
 ABUNDANT IS BACITRACIN A, A BRANCHED CYCLIC DODECAPEPTIDE. IT
 CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (ILE-CYS-LEU-D-
 GLU-ILE) WITH AN ISOLEUCINE-CYSTEINE THIAZOLINE CONDENSATION
 PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-ILE-D-PHE-
 HIS-D-ASP-ASN). IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-
 TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT
 CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,
 PHE-9, AND ASP-11).
 CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch)

 EMBL: AF007865; AAC06346.1; -.
 InterPro: IPR000873; AMP-bind.
 InterPro: IPR001242; DUF4.
 InterPro: IPR003880; Phosphopant_attach.
 Pfam: PF00501; AMP-binding; 5.
 Pfam: PF00668; Condensation; 5.
 Pfam: PF00550; pp-binding; 5.
 PRINTS: PR00154; AMPBINDING.
 PROSITE: PS00012; PHOSPHOPANTHETHEINE; 3.
 PROSITE: PS00455; AMP-BINDING; 5.
 PROSITE: PS50075; ACP-DOMAIN; 5.
 Lgase; Isomerase; Antibiotic biosynthesis; Phosphopantetheine;
 Multifunctional enzyme; Repeat.
 REPEAT 39 612 DOMAIN 1 (ISOLEUCINE-ACTIVATING).
 REPEAT 1109 1648 DOMAIN 2 (CYSTEINE-ACTIVATING).
 REPEAT 2124 2889 DOMAIN 3 (LEUCINE-ACTIVATING).
 REPEAT 3164 3732 DOMAIN 4 (GLUTAMINE-ACTIVATING).
 REPEAT 4668 5249 DOMAIN 5 (ISOLEUCINE-ACTIVATING).
 CYCLIZATION (POTENTIAL).
 REPEAT 621 1037 ACYL CARRIER (ACP) 1.
 REPEAT 544 612 ACYL CARRIER (ACP) 2.
 REPEAT 1595 1648 ACYL CARRIER (ACP) 3.
 REPEAT 2621 2889 ACYL CARRIER (ACP) 4.
 REPEAT 3664 3732 ACYL CARRIER (ACP) 5.
 REPEAT 5171 5249 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 BINDING 574 574 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 BINDING 1615 1615 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 BINDING 2651 2651 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 BINDING 3694 3694 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 BINDING 5201 5201 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 SEQUENCE 5255 AA; 598254 MW; 906E8DD68450F85B CRC64;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:51:31 ; Search time 45.07 Seconds
(without alignments)
673.585 Million cell updates/sec

Title: US-09-724-296-2

Perfect score: 4349

Sequence: 1 MTKPLTIGWKIKGLVQPTR.....KRLTARKRSRKEEVEDEK 828

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1170	26.9	218	1	GT26_SCHJA
2	984	22.6	218	1	GT27_SCHMA
3	921	21.2	220	1	GT26_SCHMA
4	710.5	16.3	220	1	GT29_FASHE
5	699	16.1	217	1	GT27_FASHE
6	681	15.7	217	1	GT28_FASHE
7	680	15.6	217	1	GT26_FASHE
8	481.5	11.1	217	1	GTM1_MOUSE
9	481.5	11.1	217	1	GTM1_RAT
10	477.5	11.0	217	1	GTMU_RABIT
11	474.5	10.9	218	1	GTM4_HUMAN
12	473.5	10.9	217	1	GTM2_HUMAN
13	473.5	10.9	217	1	GTMU_CAVPO
14	468.5	10.8	217	1	GTMU_CRILLO
15	468.5	10.8	217	1	GTMU_MOUSE
16	467.5	10.7	217	1	GTM2_MOUSE
17	465.5	10.7	217	1	GTM1_HUMAN
18	463.5	10.7	217	1	GTM5_HUMAN
19	461.5	10.6	217	1	GTM3_RAT
20	455.5	10.5	217	1	GTM2_RAT
21	452.5	10.4	217	1	GTM3_MOUSE
22	447.5	10.3	225	1	GTM3_HUMAN
23	437.5	10.1	224	1	GTM5_MOUSE
24	432.5	9.9	219	1	GTM2_CHICK
25	385.5	8.9	219	1	GTM1_DERPT
26	360	8.3	320	1	YNUJ_BACSU
27	248.5	5.7	208	1	GTP_DIRIM
28	245.5	5.6	208	1	GTP_ONCVO
29	244	5.6	209	1	GTP_MESAU
30	242	5.6	207	1	GTP_PIG
31	240	5.5	209	1	GTP_HUMAN
32	238	5.5	209	1	GTP_CRIMI
33	237	5.4	209	1	GTP2_MOUSE

34	237	5.4	209	1	GTP_RAT	P04906 rattus norv
35	236	5.4	209	1	GTP_CRILLO	P46424 cricetus
36	235	5.4	209	1	GTP_MACMU	Q28514 macaca mulla
37	233	5.4	209	1	GTP_MOUSE	P46425 mus musculus
38	230	5.3	209	1	GTP_BOVIN	P28801 bos taurus
39	222.5	5.1	223	1	GTAL_KARIT	Q08863 oryctolagus
40	219.5	5.0	208	1	GTP_CAREL	P10299 caenorhabdi
41	200.5	4.6	221	1	GTAL_MOUSE	P04903 rattus norv
42	198.5	4.6	221	1	GTAL_MOUSE	P10648 mus musculus
43	198.5	4.6	222	1	GTAL_MOUSE	P13745 mus musculus
44	194.5	4.5	221	1	GTAL_RAT	P00502 rattus norv
45	193.5	4.4	220	1	GTC2_RAT	P46418 rattus norv

ALIGNMENTS

RESULT 1	
GT26_SCHJA	
ID GT26_SCHJA	STANDARD; PRT; 218 AA.
AC P08515;	
DT 01-AUG-1988 (Rel. 08, Created)	
DT 01-AUG-1988 (Rel. 08, Last sequence update)	
DT 30-MAY-2000 (Rel. 39, Last annotation update)	
DE GLUTATHIONE S-TRANSFERASE 26 KDA (EC 2.5.1.18) (GST 26) (SJ26 ANTIGEN)	
DE (GST CLASS-ALPHA).	
OS Schistosoma japonicum (Blood fluke).	
OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian plathyhelminths;	
OC Rhabdiphora; Eulecithophora; Revertospermatia; Mediofusata;	
OC Neodermata; Trematoda; Digenea; Strigeididae; Schistosomatoidea;	
OC Schistosomatidae; Schistosoma.	
OX NCBI_TaxID=6182;	
RN [1]	
SEQUENCE FROM N.A.	
RP MEDLINE=870411520; PubMed=3095841;	
RA Smith D.B., Davern K.M., Board P.G., Tiu W.U., Garcia E.G.,	
RA Mitchell G.F.;	
RT "Mr 26,000 antigen of Schistosoma japonicum recognized by resistant	
RT WEHI 129/J mice is a parasite glutathione S-transferase.";	
RL Proc. Natl. Acad. Sci. U.S.A. 83:8703-8707(1986).	
[2]	
REVIEWS.	
RA Smith D.B., Davern K.M., Board P.G., Tiu W.U., Garcia E.G.,	
RA Mitchell G.F.;	
RL Proc. Natl. Acad. Sci. U.S.A. 84:6541-6541(1987).	
[3]	
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).	
RX MEDLINE=95276631; PubMed=7538846;	
RA Lim K., Ho J.X., Keeling K., Gilliland G.L., Ji X., Ruker F.,	
RA Carter D.C.;	
RT "Three-dimensional structure of Schistosoma japonicum glutathione S-	
RT transferase fused with a six-amino acid conserved neutralizing	
RT epitope of gp41 from HIV.";	
RL Protein Sci. 3:2233-2244(1994).	
[4]	
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).	
RX MEDLINE=95156484; PubMed=7653399;	
RA McTigue M.A., Williams D.R., Tainer J.A.;	
RT "Crystall structures of a schistosomal drug and vaccine target:	
RT glutathione S-transferase from Schistosoma japonica and its complex	
RL J. Mol. Biol. 246:21-27(1995).	
CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER	
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.	
CC -!- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE	
CC PARASITE DETOXICATION SYSTEM. OTHER FUNCTIONS ARE ALSO	
CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF	
CC HAEMATIN IN THE PARASITE GUT.	
CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE -> HX + R-S-GUTATHIONE.	
CC -!- SUBUNIT: HOMODIMER.	
CC -!- MISCELLANEOUS: THERE ARE AT LEAST TWO ISOENZYMES OF GST IN	
CC S.JAPONICUM.	
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.	

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M14654; AAB59203.1; -;
 DR PIR; A26484; A26484.
 DR PDB; 1GNE; 30-NOV-94.
 DR PDB; 1GTA; 07-FEB-95.
 DR PDB; 1GTB; 01-DEC-95.
 DR PDB; 1B8X; 12-APR-99.
 DR InterPro; IPR000521; GST.
 DR Pfam; PF00043; GST; 1.
 DR Transferase; Antigen; Multigene family; 3D-structure.
 KW SEQUENCE 218 AA; 25498 MW; 5E2AC418BD0EF13F CRC64;

Query Match 26.9%; Score 1170; DB 1; Length 218;
 Best Local Similarity 100.0%; Pred. No. 2.8e-56;
 Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 PILGYWKIKGLVQPTRLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
 DB 3 PILGYWKIKGLVQPTRLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYIDGD 62
 QY 65 VKLTQSMARIYIADKHNLGGCPKRAEISMLEGAVLDIRYGVSRISYKDFETLKVD 124
 DB 63 VKLTQSMARIYIADKHNLGGCPKRAEISMLEGAVLDIRYGVSRISYKDFETLKVD 122
 QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHVTHDFDMLYDALDVVLYMDPCLDAPFKLVCFKRR 184
 DB 123 LSKLPEMLKMFEDRLCHKTYLNGDHVTHDFDMLYDALDVVLYMDPCLDAPFKLVCFKRR 182
 QY 185 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDHPKP 220
 DB 183 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDHPKP 218

RESULT 2
 GT27_SCHMA STANDARD; PRT; 218 AA.
 ID GT27_SCHMA STANDARD; PRT; 218 AA.
 AC P35661;
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE GLUTATHIONE S-TRANSFERASE 26 KDA (EC 2.5.1.18) (GST 26) (SM26/2)
 DE ANTIGEN (GST CLASS-ALPHA).
 OS Schistosoma mansoni (Blood fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
 OC Rhabdithophora; Eulicthophora; Revertospermata; Mediofusata;
 OC Neodermata; Trematoda; Digenea; Strigeidida; Schistosomatoidea;
 OC Schistosomatidae; Schistosoma.
 OX NCBI_TaxID=6183;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92131046; PubMed=1775156;
 RA Wright M.D., Harrison R.A., Melder A.M., Newport G.R., Mitchell G.F.;
 RT "Another 26-kilodalton glutathione S-transferase of Schistosoma
 RT mansoni".
 RL Mol. Biochem. Parasitol. 49:177-179(1991).
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -!- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
 CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HAEMATIN IN THE PARASITE GUT.
 CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M73624; -; NOT_ANNOTATED_CDS.
 DR PIR; A45556; A45556.
 DR HSP; P08515; IGTB.
 DR InterPro; IPR000521; GST.
 DR Pfam; PF00043; GST; 1.
 DR Transferase; Antigen; Multigene family.
 KW SEQUENCE 218 AA; 25411 MW; D803EE9028B36185 CRC64;

Query Match 22.6%; Score 984; DB 1; Length 218;
 Best Local Similarity 82.9%; Pred. No. 2.5e-46;
 Matches 179; Conservative 18; Mismatches 19; Indels 0; Gaps 0;
 QY 5 PILGYWKIKGLVQPTRLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
 DB 3 PKLYGWKIKGLVQPTRLLEYLEEKYEELHYERDNDGVDWRNKKFELGLEFPNLPYYIDGD 62
 QY 65 VKLTQSMARIYIADKHNLGGCPKRAEISMLEGAVLDIRYGVSRISYKDFETLKVD 124
 DB 63 VKLTQSMARIYIADKHNLGGCPKRAEISMLEGAVLDIRYGVSRISYKDFETLKVD 122
 QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHVTHDFDMLYDALDVVLYMDPCLDAPFKLVCFKRR 184
 DB 123 LSKLPEMLKMFEDRLCHKTYLNGDHVTHDFDMLYDALDVVLYMDPCLDAPFKLVCFKRR 182
 QY 185 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDHPKP 220
 DB 183 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDHPKP 218

RESULT 3
 GT26_SCHMA STANDARD; PRT; 218 AA.
 ID GT26_SCHMA STANDARD; PRT; 218 AA.
 AC P15964;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE GLUTATHIONE S-TRANSFERASE 26 KDA (EC 2.5.1.18) (GST 26) (SM26/1)
 DE ANTIGEN (GST CLASS-ALPHA).
 OS Schistosoma mansoni (Blood fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
 OC Rhabdithophora; Eulicthophora; Revertospermata; Mediofusata;
 OC Neodermata; Trematoda; Digenea; Strigeidida; Schistosomatoidea;
 OC Schistosomatidae; Schistosoma.
 OX NCBI_TaxID=6183;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=PUERTO RICAN;
 RX MEDLINE=90348716; PubMed=2385266;
 RA Trottein F., Kieny M.P., Verwaerde C., Torpier G., Pierce R.J.,
 RA Balloul J.-M., Schmitt D., Lecocq J.-P., Capron A.;
 RT "Molecular cloning and tissue distribution of a 26-kilodalton
 RT glutathione S-transferase of Schistosoma japonicum and Schistosoma
 RT mansoni".
 RL Mol. Biochem. Parasitol. 41:35-44(1990).
 RN [2]
 RP SEQUENCE OF 8-218 FROM N.A.
 RC STRAIN=PUERTO RICAN;
 RX MEDLINE=90271935; PubMed=1693415;
 RA Henkle K.J., Davern K.M., Wright M.D., Ramos A.J., Mitchell G.F.;
 RT "Comparison of the cloned genes of the 26- and 28-kilodalton
 RT glutathione S-transferases of Schistosoma japonicum and Schistosoma
 RT mansoni".
 RL Mol. Biochem. Parasitol. 40:23-34(1990).
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER

CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -!- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
 CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HAEMATIN IN THE PARASITE GUT.
 CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- TISSUE SPECIFICITY: TEGUMENT AND IN SUBTEGUMENTARY PARENCHYMAL
 CC CELLS. GST 26 MAY BE ACTIVELY EXCRETED BY ADULT WORMS.
 CC -!- MISCELLANEOUS: THERE ARE AT LEAST THREE ISOENZYMES OF GST IN
 CC S.MANSONI.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M31106; AAA29888.1; -;
 CC EMBL: M26913; AAA29889.1; -;
 CC PIR: A45523; A45523.
 CC HSP: P08515; ICTB.
 CC InterPro: IPR000521; GST.
 CC Pfam: PF00043; GST; 1.
 CC Transferrase; Antigen; Multigene family.
 CC SEQUENCE 218 AA; 25401 MW; 061A6548A842D6E8 CRC64;

Query Match 21.2%; Score 921; DB 1; Length 218;
 Best Local Similarity 80.1%; Pred. No. 5.8e-43;
 Matches 173; Conservative 14; Mismatches 29; Indels 0; Gaps 0;
 QY 5 PILGWIKGLVQTRLLLEYLEEKYEHELYERDEGDKRNKKFELGLEPNNLPYYIDGD 64
 DB 3 PKGKWKYKGLVQTRLLLEHLEETEERAYDRNEIDAWSNDKFKLGLPNNLPYYIDGD 62
 QY 65 VKLTQSMATIRYIADKHNMLGCGPKRAEISMLEGAVLDIRYGVSRITAYSXKDFETLKVDF 124
 DB 63 FKLTQSMATIRYIADKHNMLGCGPKRAEISMLEGAVLDIRMGVLRITAYNKYEYETLKVDF 122
 QY 125 LSKPEMLKMPFEDRLCHTKTYLNGDHVTHPDMFLYDALDVLVYMDPMLCDAPFKLVCFKKR 184
 DB 123 LNKLPGLKMPFEDRLSNKTYLNGCVTHPDMFLYDALDVLVYMDPMLCDAPFKLVCFKK 182
 QY 185 IEATPQIDKYLKSSKYIAWPLQGWQATFGGDDHPPK 220
 DB 183 IEDLPQIKNVLNRSRYIKWPLQGWQATFGGDDTPPK 218

RESULT 4
 GT27_FASHE STANDARD; PRT; 220 AA.
 AC P56598;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE GLUTATHIONE S-TRANSFERASE 26 KDA 1 (EC 2.5.1.18) (GST1) (F01) (GST
 DE CLASS-ALPHA).
 OS Fasciola hepatica (Liver fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
 OC Rhabditophora; Eulcithophora; Revertospermata; Mediofusata;
 OC Neodermata; Trematoda; Digenea; Echinostomida; Echinostomata;
 OC Fasciolidae; Fasciolidae; Fasciola.
 OX NCBI_TaxID=6192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92155306; PubMed=1740183;
 RA Panaccio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;
 RT "Molecular characterization of cDNA sequences encoding glutathione S-
 RT transferases of Fasciola hepatica."

RL Exp. Parasitol. 74:232-237(1992).
 RN [2]
 RP ERRATUM.
 RX MEDLINE=94039664; PubMed=8224094;
 RA Panaccio M., Wilson L.R., Cramer S.L., Wijffels G.L., Spithill T.W.;
 RL Exp. Parasitol. 77:385-385(1993).
 RN [3]
 RP SEQUENCE OF 22-220 FROM N.A.
 RA Cramer S.;
 RL Patent number WO9008819, 09-AUG-1990.
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -!- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
 CC PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
 CC SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
 CC HAEMATIN IN THE PARASITE GUT.
 CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: A00993; CAA00118.1; -;
 CC InterPro: IPR000521; GST.
 CC Pfam: PF00043; Antigen; Multigene family.
 CC Transferrase; Antigen; Multigene family.
 CC INIT_MET 0 BY SIMILARITY.
 CC CONFLICT 22 22 Y -> V (IN REF. 3).
 CC CONFLICT 110 111 DP -> VS (IN REF. 3).
 CC CONFLICT 189 189 A -> P (IN REF. 3).
 CC SEQUENCE 220 AA; 25598 MW; 27B9F150B75D101F CRC64;

Query Match 16.3%; Score 710.5; DB 1; Length 220;
 Best Local Similarity 59.7%; Pred. No. 1.1e-51;
 Matches 129; Conservative 32; Mismatches 54; Indels 1; Gaps 1;
 QY 7 LGWKIKGLVQTRLLLEYLEEKYEHELYERDEGDKRNKKFELGLEPNNLPYYIDGDVK 66
 DB 4 LGWKIKGLVQTRLLLEYLEEKYEHELYERDEGDKRNKKFELGLEPNNLPYYIDDKCK 62
 QY 67 LTQSMATIRYIADKHNMLGCGPKRAEISMLEGAVLDIRYGVSRITAYSXKDFETLKVDFLS 126
 DB 63 LTQSMATIRYIADKHNMLGCGPKRAEISMLEGAVLDIRMGVLRITAYNKYEYETLKVDF 122
 QY 127 KLPEMLKMPFEDRLCHTKTYLNGDHVTHPDMFLYDALDVLVYMDPMLCDAPFKLVCFKKRIE 186
 DB 123 DLPTTKMMSDFLGKNPFLRGTSVSHVDPVMEALDAIRYLEPHCLDHPNLPQQMSRIE 182
 QY 187 AIPOIDKYLKSSKYIAWPLQGWQATFGGDDHPPKSD 222
 DB 183 ALPSKAYMESNRFKWPLNGWHAQFGGDDAPPSE 218

RESULT 5
 GT27_FASHE STANDARD; PRT; 217 AA.
 AC P31670;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE GLUTATHIONE S-TRANSFERASE 26 KDA 47 (EC 2.5.1.18) (GST47) (F047)
 DE (GST CLASS-ALPHA).
 OS Fasciola hepatica (Liver fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
 OC Rhabditophora; Eulcithophora; Revertospermata; Mediofusata;
 OC Neodermata; Trematoda; Digenea; Echinostomida; Echinostomata;

```

OC Fasciolidae; Fasciolidae; Fasciola.
OX NCBI_TaxID=6192;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92155306; PubMed=1740183;
RA Panaccio M., Wilson L.R., Crameri S.L., Wijffels G.L., Spithill T.W.;
RT "Molecular characterization of cDNA sequences encoding glutathione S-
transferases of Fasciola hepatica.";
RL Exp. Parasitol. 74:232-237(1992).
RN [2]
RP ERRATUM.
RX MEDLINE=94039664; PubMed=8224094;
RA Panaccio M., Wilson L.R., Crameri S.L., Wijffels G.L., Spithill T.W.;
RL Exp. Parasitol. 77:385-385(1993).
RN [3]
RP SEQUENCE OF 7-105 FROM N.A.
RX Crameri S.;
RA Patent number WO9008819, 09-AUG-1990.
RN [4]
RP X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
RX Rosjohn J., Feil S.C., Wilce M.C.J., Sexton J.L., Spithill T.W.,
RA Parker M.W.;
RT "Crystallization, structural determination and analysis of a novel
parasitic vaccine candidate: Fasciola hepatica glutathione
S-transferase.";
RL J. Mol. Biol. 273:857-872(1997).
CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -!- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
HAEMATIN IN THE PARASITE GUT.
CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE -> HX + R-S-GLUTATHIONE.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL: M77681; AAA29140.1; -.
DR EMBL: A00996; CAA00121.1; -.
DR PDB: 1PHE; 29-JUL-98.
DR InterPro: IPR000521; GST.
DR Pfam: PF00043; GST; 1.
KW Transferase; Antigen; Multigene family; 3D-structure.
FT INIT_MET 0
FT CONFLICT 65 65 T -> I (IN REF. 3).
FT CONFLICT 102 105 RTGF -> FEEL (IN REF. 3).
SQ SEQUENCE 217 AA; 25281 MW; 0FB8BBFE63029E03 CRC64;

Query Match 16.1%; Score 699; DB 1; Length 217;
Best Local Similarity 58.2%; Pred. No. 4.3e-31;
Matches 124; Conservative 36; Mismatches 53; Indels 0; Gaps 0;

OY 7 LGYWKIKGLVQPTKLLLEYLEEKYEHEHYERDEGDKWRNKKFELGFPNLPYYIDGDKV 66
DB 4 LGYWKIKGLVQPTKLLLEYLEEKYEHEHYERDEGDKWRNKKFELGFPNLPYYIDGDKV 63

OY 67 LQTSMAIRYIADKHNKMLGCGPKEAETSMLEGAVIDRIGVSRISYKDFETLKVDFLS 126
DB 64 LQTSVAMRYIADKHNKMLGCTTTEPARARISMEGAANDLRIGRVCYNPKFEVKEEYVK 123

OY 127 KLPEMLKMFEDRLCHTKYTLNGDHVTHPDPMYDALDVVLYMDPMCLDAPPKLVCFKKRIE 186
DB 124 ELPEKTLKMSDFLGRHYLTGSSVSHVDFMELYMTLDSYRLAPHCLDFPPLKFKRSRIE 183

```

```

OY 187 AIPOLDKYLKSSKYIAWPLQGWQOATEGGGHHPP 219
DB 184 ALPKIKAYMESKRFJKWPLNGWAASFAGDAPP 216

```

```

RESULT 6
GT28_FASHE STANDARD; PRT; 217 AA.
AC P31671;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUTATHIONE S-TRANSFERASE 26 KUA 7 (EC 2.5.1.18) (GST7) (FH7) (GST
DE CLASS-ALPHA).
OS Fasciola hepatica (Liver fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
OC Rhabditophora; Eulicthophora; Revertospermata; Mediofusata;
OC Neodermata; Trematoda; Digenea; Echinostomida; Echinostomata;
OC Fasciolidae; Fasciolidae; Fasciola.
OX NCBI_TaxID=6192;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92155306; PubMed=1740183;
RA Panaccio M., Wilson L.R., Crameri S.L., Wijffels G.L., Spithill T.W.;
RT "Molecular characterization of cDNA sequences encoding glutathione S-
transferases of Fasciola hepatica.";
RL Exp. Parasitol. 74:232-237(1992).
RN [2]
RP ERRATUM.
RX MEDLINE=94039664; PubMed=8224094;
RA Panaccio M., Wilson L.R., Crameri S.L., Wijffels G.L., Spithill T.W.;
RL Exp. Parasitol. 77:385-385(1993).
RN [3]
RP SEQUENCE OF 8-217 FROM N.A.
RA Crameri S.;
RA Patent number WO9008819, 09-AUG-1990.
CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -!- FUNCTION: GST ISOENZYMES APPEAR TO PLAY A CENTRAL ROLE IN THE
PARASITE DETOXIFICATION SYSTEM. OTHER FUNCTIONS ARE ALSO
SUSPECTED INCLUDING A ROLE IN INCREASING THE SOLUBILITY OF
HAEMATIN IN THE PARASITE GUT.
CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE -> HX + R-S-GLUTATHIONE.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ALPHA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL: M77680; AAA29139.1; -.
DR EMBL: A00994; CAA00119.1; -.
DR HSP; P08515; IGTB.
DR InterPro: IPR000521; GST.
DR Pfam: PF00043; GST; 1.
KW Transferase; Antigen; Multigene family.
FT INIT_MET 0
FT CONFLICT 35 42 NOREKWLIG -> MIGENWHA (IN REF. 3).
FT CONFLICT 188 196 IKEMKSR -> SRYMSRA (IN REF. 3).
SQ SEQUENCE 217 AA; 25196 MW; 0099E1F59E49A49E CRC64;

Query Match 15.7%; Score 681; DB 1; Length 217;
Best Local Similarity 58.7%; Pred. No. 4e-30;
Matches 125; Conservative 29; Mismatches 59; Indels 0; Gaps 0;

OY 7 LGYWKIKGLVQPTKLLLEYLEEKYEHEHYERDEGDKWRNKKFELGFPNLPYYIDGDKV 66
DB 124 ELPEKTLKMSDFLGRHYLTGSSVSHVDFMELYMTLDSYRLAPHCLDFPPLKFKRSRIE 183

```



```

RP PRELIMINARY SEQUENCE OF 1-24.
RX MEDLINE-86042634; PubMed-3864155;
RA Mannervik B., Alin P., Guthenberg C., Jansson H., Thair M.K.,
RA Warholm M., Joernvall H.;
RT "Identification of three classes of cytosolic glutathione transferase
RT common to several mammalian species: correlation between structural
RT data and enzymatic properties.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:7202-7206(1985).
RN [6]
RP CHARACTERIZATION.
RC STRAIN-CD-1; TISSUE=Liver;
RX MEDLINE-96189427; PubMed-8605288;
RA Mitchell A.E., Morin D., Lane M.W., Jones A.D.;
RT "Purification, mass spectrometric characterization, and covalent
RT modification of murine glutathione S-transferases.";
RL Chem. Res. Toxicol. 8:1054-1062(1995).
CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- MASS SPECTROMETRY: MW=25838.4; MW_ERR=2; METHOD=ELECTROSPRAY.
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; J03952; AAA37747.1; -.
DR EMBL; J04632; AAA37705.1; -.
DR EMBL; L13448; -. NOT_ANNOTATED_CDS.
DR PIR; A20831; A20831.
DR PIR; A28946; A28946.
DR PIR; I24735; I24735.
DR PIR; A34159; A34159.
DR PIR; S33860; S33860.
DR HSSP; P04905; 6GSY.
DR SWISS-2DPAGE; P10649; MOUSE.
DR MGD; MGI:95860; Gstml.
DR InterPro; IPR000521; GST.
DR InterPro; IPR003081; GST_mu.
DR Pfam; PF00043; GST; 1.
DR PRINTS; PR01267; GSTRNSFRASEM.
KW Transferase; Multigene family.
FT INIT_MET 0
SQ SEQUENCE 217 AA; 25839 MW; A1EE3938F590B829 CRC64;

```

```

RESULT 9
GTM1_RAT
ID GTM1_RAT STANDARD; PRT; 217 AA.
AC P04905;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLUTATHIONE S-TRANSFERASE YB1 (EC 2.5.1.18) (CHAIN 3) (GST M1-1)
DE (GST CLASS-MU).
DE GSTM1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (CLONE PGTR200).
RX MEDLINE-86312882; PubMed-2875437;
RA Lai H.-C.J., Grove G., Tu C.-P.D.;
RT "Cloning and sequence analysis of a cDNA for a rat liver glutathione
RT S-transferase Yb subunit.";
RL Nucleic Acids Res. 14:6101-6114(1986).
RN [2]
RP SEQUENCE FROM N.A. (CLONE PGTA/C44).
RX MEDLINE-86033768; PubMed-3840477;
RA Ding G.-J.-F., Lu A.Y.H., Pickett C.B.;
RT "Rat liver glutathione S-transferases. Nucleotide sequence analysis
RT of a Yb1 cDNA clone and prediction of the complete amino acid
RT sequence of the Yb1 subunit.";
RL J. Biol. Chem. 260:13268-13271(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-86224097; PubMed-3011803;
RA Ding G.-J.-F., Ding V.D.-H., Rodkey J.A., Bennett C.D., Lu A.Y.H.,
RA Pickett C.B.;
RT "Rat liver glutathione S-transferases. DNA sequence analysis of a Yb2
RT cDNA clone and regulation of the Yb1 and Yb2 mRNAs by
RT phenobarbital.";
RL J. Biol. Chem. 261:7952-7957(1986).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE-87308179; PubMed-3040722;
RA Chang C., Saltzman A.G., Sorensen N.S., Hilpakka R.A., Liao S.;
RT "Identification of glutathione S-transferase Yb1 mRNA as the
RT androgen-repressed mRNA by cDNA cloning and sequence analysis.";
RL J. Biol. Chem. 262:11901-11903(1987).
RN [5]
RP SEQUENCE OF 1-23.
RC STRAIN=Wistar; TISSUE=Olfactory epithelium;
RX MEDLINE-93277499; PubMed-8503873;
RA Ben-Arie N., Khen M., Lancet D.;
RT "Glutathione S-transferases in rat olfactory epithelium:
RT purification, molecular properties and odorant biotransformation.";
RL Biochem. J. 292:379-384(1993).
RN [6]
RP MUTAGENESIS OF CYS-86.
RX MEDLINE-91354218; PubMed-1883338;
RA Hsieh J.-C., Huang S.-C., Chen W.-L., Lai Y.-C., Tam M.F.;
RT "Cysteine-86 is not needed for the enzymic activity of glutathione S-
RT transferase 3-3.";
RL Biochem. J. 278:293-297(1991).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE-93041702; PubMed-1420139;
RA Ji X., Zhang P., Armstrong R.N., Gilliland G.L.;
RT "The three-dimensional structure of a glutathione S-transferase from
RT the mu gene class. Structural analysis of the binary complex of
RT isoenzyme 3-3 and glutathione at 2.2-A resolution.";
RL Biochemistry 31:10169-10184(1992).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RA Fu J.-H., Rose J., Tam M.F., Wang B.-C.;
RT "New crystal forms of a mu-class glutathione S-transferase from rat
RT liver.";

```

Acta Crystallogr. D 50:219-224(1994).

[9]

RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).

RX MEDLINE=9415386; PubMed=8110735;

JI X., Johnson W.W., Sesay M.A., Dickert L., Prasad S.M., Ammon H.L.,

RA Armstrong R.N., Gilliland G.L.;

RT "Structure and function of the xenobiotic substrate binding site of a

RT glutathione S-transferase as revealed by X-ray crystallographic

RT analysis of product complexes with the diastereomers of 9-(S-

RT glutathionyl)-10-hydroxy-9,10-dihydrophenanthrene.";

RL Biochemistry 33:1043-1052(1994).

CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER

CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.

CC -!- FUNCTION: THE OLFACTORY GST MAY BE CRUCIAL FOR THE ACUITY OF THE

CC OLFACTORY PROCESS.

CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.

CC -!- SUBUNIT: HOMODIMER OR HETERODIMER.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC -!- MISCELLANEOUS: RAT LIVER GLUTATHIONE S-TRANSFERASES (GST) CONSIST

CC OF BINARY COMBINATIONS OF THREE MAJOR CLASSES OF SUBUNITS

CC DESIGNATED: YA (25 KDA), YB (27 KDA) AND YC (28 KDA).

CC -!- MISCELLANEOUS: YB SUBCLASS SELECTIVELY BINDS STEROID HORMONES.

CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

EMBL; X04229; CAA27811.1; -

DR EMBL; M1719; AAA41287.1; -

DR EMBL; J02810; AAA41293.1; -

DR PIR; A24085; A24085.

DR PIR; A25510; A25510.

DR PIR; A29794; A29794.

DR PIR; S17167; S17167.

DR PDB; 1GSB; 31-OCT-93.

DR PDB; 1GSC; 31-OCT-93.

DR PDB; 2GST; 31-OCT-93.

DR PDB; 3GST; 31-JAN-94.

DR PDB; 4GST; 31-OCT-93.

DR PDB; 5GST; 31-OCT-93.

DR PDB; 6GST; 08-NOV-96.

DR PDB; 6GSU; 08-NOV-96.

DR PDB; 6GSV; 08-NOV-96.

DR PDB; 6GSW; 08-NOV-96.

DR PDB; 6GSX; 08-NOV-96.

DR PDB; 6GSY; 08-NOV-96.

DR PDB; 5FWG; 27-JAN-99.

DR InterPro; IPR000521; GST.

DR InterPro; IPR003081; GST_mu.

DR Pfam; PF00043; GST; 1.

DR PRINTS; PR01267; GSTRNSFRASEM.

DR Transferase; Multigene family; 3D-structure; Olfaction.

FT INIT_MET 0

FT MUTAGEN 86

FT CONFLICT 168

FT CONFLICT 198

FT STRAND 2

FT TURN 11

FT TURN 13

FT HELIX 14

FT TURN 23

FT STRAND 27

FT TURN 37

FT TURN 43

FT HELIX 46

FT TURN 47

FT TURN 50

FT STRAND 61

FT TURN 65

FT STRAND 67

FT STRAND 70

FT HELIX 72

CC C->S: NO CHANGE IN ACTIVITY.

CC I -> N (IN REF. 3).

CC KS -> NC (IN REF. 2).

FT TURN 83 84

FT HELIX 90 114

FT TURN 115 118

FT TURN 117 118

FT HELIX 119 128

FT TURN 129 129

FT HELIX 130 141

FT TURN 142 143

FT TURN 154 154

FT HELIX 155 169

FT TURN 171 176

FT HELIX 178 188

FT TURN 189 189

FT HELIX 191 196

FT TURN 197 198

FT TURN 200 201

FT TURN 210 211

SQ SEQUENCE 217 AA; 25782 MW; 2ACE8D49DA785118 CRC64;

Query Match 11.1%; Score 481.5; DB 1; Length 217;

Best Local Similarity 44.4%; Pred. No. 1.9e-19;

Matches 91; Conservative 38; Mismatches 71; Indels 5; Gaps 1;

QY 6 ILGYWKIKGIIVOPTRLLLEVEEYEEHLYERDEG-----DKWRNKKPFELGFEFNLPLY 60

DB 3 ILGYWVRGLTIPRLLETTSSYEKKRYAMGDADPDYDRSOWLNKFKLGLDFNLPLYL 62

QY 61 IDGDVKLTOSMAITRYIADKHNNMLGCGPKERAEISMLEGAVLDIRYGVSRVAYSKDFETL 120

DB 63 IDGSRKKTQSNAMRYLARKHKLHLCGETEERIRADIVENQVMDNRMLIMLCYNPDFEQK 122

QY 121 KVDFLSKPEMLKMFEDRLCHTKTYLNGDQIVTHPDMFLYDALDVLVYMDPCLDAFPKLV 180

DB 123 KPEFLKTIPEKNKLYSEFLGKRPWFAGDKVTYVDFLAYDILDQYHIFEPKCLDAFPNLKD 182

QY 181 EKKRIEAIPOIDKYLKSSKYIAMPL 205

DB 183 FLARFELKRIKISAYMKSSRYLSTPI 207

RESULT 10

ID GTMUL_RABIT STANDARD; PRT; 217 AA.

AC P46409;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE GLUTATHIONE S-TRANSFERASE MU 1 (EC 2.5.1.18) (GST MU 1) (GST CLASS-

DE MU).

OS Oryctolagus cuniculus (Rabbit).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI_TaxID=9986;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=95251394; PubMed=7733673;

RA Lee S.H., Lee S.H., Han J.S., Kim Y.S., Koh J.K.;

RT "Cloning and expression of a cDNA for mu-class glutathione

RT S-transferase from rabbit liver.";

RL Arch. Biochem. Biophys. 318:424-429(1995).

CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER

CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.

CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.

CC -!- SUBUNIT: HOMODIMER.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC -!- TISSUE SPECIFICITY: WELL EXPRESSED IN RABBIT LIVER, BRAIN, AND

CC KIDNEY.

CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91239584; PubMed=2034681;
 RA Vorachek W.R., Pearson W.R., Rule G.S.;
 RT "Cloning, expression, and characterization of a class-mu glutathione
 RT transferase from human muscle, the product of the GST4 locus.";
 RP Proc. Natl. Acad. Sci. U.S.A. 88:4443-4447(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).
 RX MEDLINE=94238693; PubMed=8182750;
 RA Raghunathan S., Chandross R.J., Kretsinger R.H., Allison T.J.,
 RA Penington C.J., Rule G.S.;
 RT "Crystal structure of human class mu glutathione transferase GSTM2-2.
 RT Effects of lattice packing on conformational heterogeneity.";
 RL J. Mol. Biol. 238:815-832(1994).
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- TISSUE SPECIFICITY: THIS IS A MUSCLE ISOZYME.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M63509; AAA60963.1; -
 DR PIR: A39375; A39375.
 DR PDB: 1HNA; 31-JAN-94.
 DR PDB: 1HNB; 31-JAN-94.
 DR PDB: 1HNC; 31-JAN-94.
 DR PDB: 2GTU; 02-MAR-99.
 DR MIM: 138380;
 DR InterPro: IPR000521; GST.
 DR InterPro: IPR003081; GST_mu.
 DR Pfam: PF00043; GST; 1.
 DR PRINTS: PR01267; GSTRNSFRASEM.
 DR Transferase; Multigene family; 3D-structure.
 FT INIT_MET 0 0
 FT STRAND 2 5
 FT STRAND 7 7
 FT TURN 11 12
 FT HELIX 13 22
 FT TURN 23 24
 FT STRAND 27 29
 FT STRAND 32 32
 FT STRAND 35 35
 FT TURN 38 39
 FT STRAND 41 41
 FT HELIX 43 49
 FT TURN 50 52
 FT STRAND 61 64
 FT TURN 65 66
 FT STRAND 67 70
 FT HELIX 72 81
 FT TURN 82 84
 FT HELIX 90 113
 FT TURN 114 115
 FT TURN 117 118
 FT TURN 119 127
 FT TURN 128 129
 FT HELIX 130 141
 FT TURN 142 143
 FT TURN 146 146
 FT STRAND 146 146

FT TURN 147 147
 FT STRAND 148 149
 FT TURN 150 150
 FT STRAND 151 151
 FT HELIX 154 169
 FT HELIX 171 176
 FT HELIX 178 189
 FT TURN 191 197
 FT TURN 198 198
 FT TURN 214 215
 SQ SEQUENCE 217 AA; 25613 MW; 25603A909482CA39 CRC64;
 Query Match 10.9%; Score 473.5; DB 1; Length 217;
 Best Local Similarity 44.3%; Pred. No. 5e-19;
 Matches 94; Conservative 35; Mismatches 78; Indels 5; Gaps 1;
 QY 7 LGYWKIKGLVQPTRLLEYLEEKYEHLRYERDEG-----DKWRNKKFELGLEFPNLPYYI 61
 DB 4 LGYWNIRGLAHSIRLLEYSYEEKKYYTMGDAPDYDRSQWLNEKPKGLGLDFNLPYLI 63
 QY 62 DGDVKLTOSMAIIRYIADKHNMLGCGCPKERAIEISMLECAVLDIRYGSRIAYSKDFTLK 121
 DB 64 DQTHKITQSNAILRYIARKHNLCGESEKEQIREIDILENQFMDSRMQLAKCYDPDFEKLK 123
 QY 122 VDFLSKLPMLKMFEDRLCHTYLNGDHVTHPDFMPLYDALDWVLYMDPMCLDAFPKLVCF 181
 DB 124 PEYLALEPEMLKLSQFGLGKOPFWLGDKITFVDFLAVDLVLRNOVFEPSCLDAPNKLKF 183
 QY 182 KRIEAIPOIDKYIKSSKYIAMIPOGWOATFG 213
 DB 184 ISRPEGLEKISAYMKSSRFLPRPVTKMAVWG 215
 RESULT 13
 GTMU_CAVPO
 ID GTMU_CAVPO STANDARD; PRT; 217 AA.
 AC P16413;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE GLUTATHIONE S-TRANSFERASE B (EC 2.5.1.18) (GST B) (GST CLASS-MU).
 GN GSTM1.
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Hystriognathi; Caviidae; Cavia.
 OX NCBI_TaxID=10141;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Liver;
 RX MEDLINE=90236961; PubMed=2332413;
 RA Kamei K., Oshino R., Hara S.;
 RT "Amino acid sequence of glutathione S-transferase b from guinea pig
 RT liver.";
 RL J. Biochem. 107:111-117(1990).
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
 DR PIR: JX0095; JX0095.
 DR HSSP: P04905; 6GSV.
 DR InterPro: IPR000521; GST.
 DR InterPro: IPR003081; GST_mu.
 DR Pfam: PF00043; GST; 1.
 DR PRINTS: PR01267; GSTRNSFRASEM.
 KW Transferase; Multigene family.
 SQ SEQUENCE 217 AA; 25719 MW; D29F7951D4E9365E CRC64;

Query Match 10.9%; Score 473.5; DB 1; Length 217;
 Best Local Similarity 45.1%; Pred. No. 5e-19;

Matches 92; Conservative 34; Mismatches 73; Indels 5; Gaps 1;

QY 7 LGYWKIKGLVPTRLLEYLEEKEHYERDEG-----DKWRNKKFELGLEFPNLPYYI 61
 DB 4 LGYWNIRGLTHPIRLILEYTSYGEYKRYNMGDAPDYSOWLNEKFKGLGDFPNLPYLI 63
 QY 62 DGDVKTOSMAIRYIADKHNMGLGCPKRAEISMLEGAVLDIRYGVSRISYKDFETLK 121
 DB 64 DGTHTKLTOSNAILRYIARKHNLGCGTEETIRMDILENOVMDIRMLIMLCYSPDFEOKK 123
 QY 122 VDFLSKLPKMLKMPEDRLCHKTYLNGDHVTHPDMFLYDALDVLVYMDPMDLADAPPKLVCF 181
 DB 124 AEFLEGIPDKMKLFSQFLGKLPWFAGNKLTYVDLAYDVLVDYOYRMLPKCLEAPNPKDF 183
 QY 182 KRIEAIPOIDKYLKSSKYIAWPL 205
 DB 184 ISRFEGLEKISSYKSSRFLPKPL 207

RESULT 14
 GTMU_CRILLO STANDARD; PRT; 217 AA.
 AC Q00285;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE GLUTATHIONE S-TRANSFERASE Y1 (EC 2.5.1.18) (CHAIN 3) (GST CLASS-MU).
 OS Cricetus longicaudatus (Long-tailed hamster) (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Cricetulus.
 OC NCBI_TaxID=10030;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91031445; PubMed=2226437;
 RA de Saint Vincent B.R., Hyrien O., Debatisse M., Buttin G.;
 RT "Complication of mu class glutathione S-transferase genes and an
 RT adenylate deaminase gene in cotormycin-resistant Chinese hamster
 RT fibroblasts";
 RL Eur. J. Biochem. 193;19-24(1990).
 CC
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X57489; CAA40726.1; -;
 CC PIR; S13202; S13202.
 CC HSSP; P04905; 6GST.
 CC InterPro; IPR000521; GST.
 CC DR InterPro; IPR003081; GST_mu.
 CC DR Pfam; PF00043; GST; 1.
 CC DR PRINTS; PR01267; GSTRNSFRASEM.
 CC TRANSFERASE; Multigene family.
 FT INIT MET 0 BY SIMILARITY.
 SQ SEQUENCE 217 AA; 2568 MW; 34D8A8FBD8627C0 CRC64;

Query Match
 Best Local Similarity 10.8%; Score 468.5; DB 1; Length 217;
 Matches 89; Conservative 36; Mismatches 75; Indels 5; Gaps 1;

QY 6 ILGYWKIKGLVPTRLLEYLEEKEHYERDEG-----DKWRNKKFELGLEFPNLPYYI 60
 DB 4 LGYWNIRGLTHPIRLILEYTSYGEYKRYNMGDAPDYSOWLNEKFKGLGDFPNLPYLI 63

Db 3 ILGYWVRGLTPIRLILLEYTSSYEKKYTMGDAPDSDRSOWLNEKFKGLGDFPNLPYL 62
 QY 61 IGDVKTOSMAIRYIADKHNMGLGCPKRAEISMLEGAVLDIRYGVSRISYKDFETLK 120
 DB 63 IGDSHKLTOSNAILRYIARKHNLGCGTEETIRMDILENOVMDIRMLIMLCYNDFEOKK 122
 QY 121 KYDFLSKLPKMLKMPEDRLCHKTYLNGDHVTHPDMFLYDALDVLVYMDPMDLADAPPKLVCF 180
 DB 123 KPEFLKTIPEKWKMYSEFLGKLPWFAGNKLTYVDLAYDVLVDYOYRMLPKCLEAPNPKDF 182
 QY 181 FKRIEAIPOIDKYLKSSKYIAWPL 205
 DB 183 FLAREGLEKISSYKSSRFLPKPL 207

RESULT 15
 GTMU_MESAU STANDARD; PRT; 217 AA.
 AC P30116;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE GLUTATHIONE S-TRANSFERASE (EC 2.5.1.18) (GST CLASS-MU).
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OC NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Smooth muscle;
 RX MEDLINE=92335246; PubMed=1631097;
 RA Fan W.M., Trifiletti R., Norris J.S., Cooper T.M.;
 RT "Cloning of a mu-class glutathione S-transferase gene and
 RT identification of the glucocorticoid regulatory domains in its 5'
 RT flanking sequence";
 RL Proc. Natl. Acad. Sci. U.S.A. 89;6104-6108(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Smooth muscle;
 RX MEDLINE=92049380; PubMed=1944302;
 RA Norris J.S., Schwartz D.A., MacLeod S.L., Fan W.M., O'Brien T.J.,
 RA Harris S.H., Trifiletti R., Cornett L.E., Cooper T.M., Levi W.M.,
 RA Smith K.G.;
 RT "Cloning of a mu-class glutathione S-transferase complementary DNA
 RT and characterization of its glucocorticoid inducibility in a smooth
 RT muscle tumor cell line";
 RL Mol. Endocrinol. 5;979-986(1991).
 CC
 CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
 CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
 CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-GLUTATHIONE.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. MU FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M59772; AAA37075.1; -;
 CC EMBL; X61033; CAA43368.1; -;
 CC PIR; A23732; A23732.
 CC HSSP; P04905; 6GST.
 CC InterPro; IPR000521; GST.
 CC DR InterPro; IPR003081; GST_mu.
 CC DR Pfam; PF00043; GST; 1.
 CC DR PRINTS; PR01267; GSTRNSFRASEM.
 CC TRANSFERASE; Multigene family.
 FT INIT MET 0 BY SIMILARITY.
 SQ SEQUENCE 217 AA; 2568 MW; 34D8A8FBD8627C0 CRC64;

```

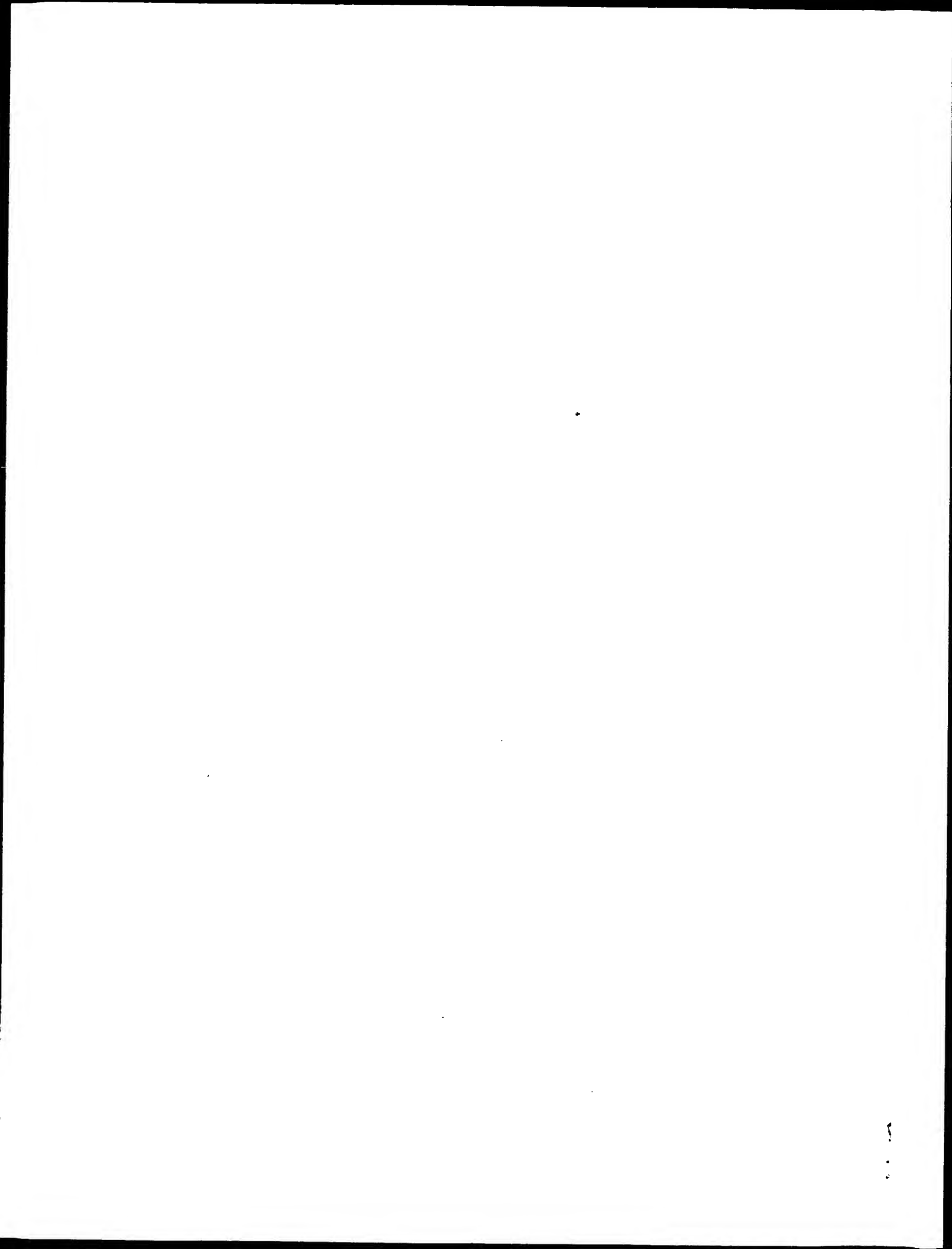
FT INIT MET 0 0 BY SIMILARITY.
SQ SEQUENCE 217 AA: 25558 MW: C59812A4A30B6375 CRC64:

Query Match 10.8%; Score 468.5; DB 1; Length 217;
Best Local Similarity 44.7%; Pred No. 9.2e-19;
Matches 89; Conservative 37; Mismatches 68; Indels 5; Gaps 1;

QY 7 LGYWKIKGLVOPTRLLLEYLEEKYEELHYERDEG-----DKWRNKKFELGLEFPNLPYI 61
Db 4 LGYWDIRGLAHAIIRLLLEVTDTSYEEKYTMGDAPNFDRSQWLNEKFKLGLDLPNLPYLI 63
QY 62 DGDVKLTQSMIIIRYIADKHNMLGGCPKERAETISMLEGAVLDIRYGVSRVAYSKDFETLK 121
Db 64 DGSFKITQSNAILRYIARKHDLGCGETEERKIQLDIENQAMDTROMOLAMVCYSPDFEKKR 123
QY 122 VDFLSKLPEMKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCF 181
Db 124 PEYLEGLPEKMKLYSEFLGKRSWFAGDKITYVDFLIYDVLQDHRIFAPKCLDAFPNPKDF 183
QY 182 KKRTEAIPQIDKYLKSSKY 200
Db 184 LARFEGCLKISDYMKSSRF 202

```

Search completed: January 15, 2002, 14:07:00
Job time: 929 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:49:59 ; Search time 120.01 Seconds
(without alignments)
228.991 Million cell updates/sec

Title: US-09-724-296-4
Perfect score: 179
Sequence: 1 DDHAPREMFCDLXPPIWRG.....KRLTARKRRSRKEVEDEK 371

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: .522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1979	100.0	371	21	AA144499 S. pombe delta228-
2	1979	100.0	600	21	AA144500 GST signal peptide
3	1979	100.0	828	21	AA144498 GST signal peptide
4	960	48.5	656	17	AAW01618 Neurospora crassa
5	910	46.0	626	21	AA144502 N. crassa delta228
6	910	46.0	626	21	AA144503 B. subtilis delta2
7	230.5	11.6	294	21	AA144505 D. radiodurans del
8	108.5	5.5	388	21	AA144507 Arabidopsis thalia
9	108.5	5.5	470	21	AA144509 Arabidopsis thalia
10	108.5	5.5	486	21	AA144511 Arabidopsis thalia
11	107.5	5.4	615	21	AA18697 Arabidopsis thalia

12	107.5	5.4	615	21	AA18697 Arabidopsis thalia
13	107.5	5.4	633	21	AA18696 Arabidopsis thalia
14	107.5	5.4	633	21	AA18696 Arabidopsis thalia
15	107.5	5.4	662	21	AA18695 Arabidopsis thalia
16	107.5	5.4	662	21	AA18695 Arabidopsis thalia
17	104	5.3	463	18	AAW34262 A beta subunit of
18	101	5.1	800	11	AA18696 Arabidopsis thalia
19	99	5.0	463	21	AA18696 Arabidopsis thalia
20	99	5.0	471	21	AA18696 Arabidopsis thalia
21	99	5.0	480	21	AA18696 Arabidopsis thalia
22	98.5	5.0	321	21	AA18696 Arabidopsis thalia
23	98.5	5.0	337	21	AA18696 Arabidopsis thalia
24	98.5	5.0	338	21	AA18696 Arabidopsis thalia
25	97.5	4.9	470	21	AA18696 Arabidopsis thalia
26	97.5	4.9	478	21	AA18696 Arabidopsis thalia
27	97.5	4.9	487	21	AA18696 Arabidopsis thalia
28	97	4.9	533	21	AA18696 Arabidopsis thalia
29	94	4.7	2110	21	AA18696 Arabidopsis thalia
30	93.5	4.7	2048	22	AA18696 Arabidopsis thalia
31	93	4.7	329	18	AA18696 Arabidopsis thalia
32	93	4.7	605	22	AA18696 Arabidopsis thalia
33	93	4.7	747	21	AA18696 Arabidopsis thalia
34	93	4.7	747	21	AA18696 Arabidopsis thalia
35	92.5	4.7	1059	21	AA18696 Arabidopsis thalia
36	91.5	4.6	653	19	AA18696 Arabidopsis thalia
37	91.5	4.6	1128	20	AA18696 Arabidopsis thalia
38	90.5	4.6	413	22	AA18696 Arabidopsis thalia
39	90	4.5	677	20	AA18696 Arabidopsis thalia
40	90	4.5	677	22	AA18696 Arabidopsis thalia
41	90	4.5	793	18	AA18696 Arabidopsis thalia
42	89.5	4.5	349	21	AA18696 Arabidopsis thalia
43	89.5	4.5	1780	22	AA18696 Arabidopsis thalia
44	89.5	4.5	3066	18	AA18696 Arabidopsis thalia
45	89	4.5	215	21	AA18696 Arabidopsis thalia

ALIGNMENTS

RESULT 1
AA144499
ID AAY44499 standard; Protein; 371 AA.
XX
XX AAY44499;
XX
XX 27-MAR-2000 (first entry)
DT
XX
DE S. pombe delta228-UV damage endonuclease.
XX
XX Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
KW glutathione-S-transferase signal peptide; uvel+ gene product;
KW irradiation; DNA damage; UV radiation damage; photoproduct;
KW abasic site; platinum diadduct; mismatched nucleotide pairing;
KW nucleotide alkylation; skin cancer.
XX
XX Schizosaccharomyces pombe.
XX
XX WO963828-A1.
XX
XX 16-DEC-1999.
PD
XX
XX 08-JUN-1999; 99WO-US12910.
PF
XX
XX 08-JUN-1998; 98US-0088521.
PR
XX
XX 18-MAY-1999; 99US-0134752.
PA
XX
XX (UYEM-) UNIV EMORY.
XX
XX Doetsch PW, Kaur B, Avery AM;
XX
XX WPI: 2000-116417/10.
DR
XX
XX N-PSDB: AA229859.
XX

PT A new truncated ultraviolet damage endonuclease for treatment of skin
 XX cancers -
 PS Claim 13; Page 53; 133pp; English.
 XX
 CC The present sequence is delta228-UV damage endonuclease. Delta228-UVDE is
 CC a 288 amino acid deletion of the N-terminal of the S. pombe uvel+ gene
 CC product. This is expressed in frame with a GST leader sequence to
 CC generate a fusion protein. This provides stable endonuclease fragments
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, platinum diadduct, an intercalated
 CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 SQ Sequence 371 AA;

Query Match 100.0%; Score 1979; DB 21; Length 371;
 Best Local Similarity 100.0%; Pred. No. 2.3e-187;
 Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDHAPREMFDCLDKPIPWGRGLGYACLTILSRMKERVFCSTCRITTIQDGLSVKQL 60
 Db 1 ddhapremfdcldkpwwgrgrlgyacintilrsmkervfcsrtcrittiqdglsvkql 60
 QY 61 GTQNVLDLILKLVENHNFGIHFMRVSSDLFPFASHAKYGYTLFQAQSHLEEVGKLANKYN 120
 Db 61 gtqnvldlilkvewnhnfgihfmrsvssdlfpfashakgytlfqaqshleevgklankyn 120
 QY 121 HRLTMHPGQYTOIASPREVVVDSATRDLAYHDEILSRMKLNQKNDVLIHILGGTFEG 180
 Db 121 hrltmhpggytqiasprevvvdsairdlayhdeilsrmlneqinkndavliihlggtfeg 180
 QY 181 KKETLDRFKNTQRLSDSVKARLVLENDVSVSDLLPLCOELNPLVLDWHHNIVPG 240
 Db 181 kketldrfknyqrlsdsvkarlvlendvsvsqdllplcqelnplvldwhhnivpg 240
 QY 241 TLREGSLDMLPIPTIRETWKGTQKOHYSESADPTAISGMKRRASDRVDFPFCDP 300
 Db 241 tlregslldmpliptiretwrkgtqkqhyssadptaisgmkrasdrvfdfppcdp 300
 QY 301 TMDLMEAKEKEQAVFELCRRYELONPCPLEIMGPEYDQTRDGYPPGAEKRLTARKRR 360
 Db 301 tmdlmeakekeqavfelcrryelnppcpleimgpeydtqrdgyppgaekrltarkrr 360
 QY 361 SRKEEVEDEK 371
 Db 361 srkeeveedek 371

RESULT 2
 ID AA144500 standard; Protein; 600 AA.
 XX
 AC AA144500;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE GST signal peptide and delta228 S. pombe UVDE fusion protein.
 XX
 KW GST signal peptide; glutathione-S-transferase signal peptide;
 KW ultraviolet damage endonuclease; UVDE; delta228-UVDE;
 KW repair-deficient E. coli strain; UV irradiation; DNA damage;
 KW UV radiation damage; photoproduct; abasic site; platinum diadduct;
 KW mismatched nucleotide pairing; nucleotide alkylation;
 KW fusion protein; skin cancer.
 XX
 OS Schizosaccharomyces pombe.
 OS Synthetic.
 XX
 PH Key Location/Qualifiers

FT Misc-difference 11 /note= "Encoded by aaaa"
 FT Misc-difference 85 /note= "Encoded by gtt"
 FT Misc-difference 147 /note= "Encoded by gtt"
 FT
 XX W09963828-A1.
 XX
 PD 16-DEC-1999.
 XX
 XX 08-JUN-1999; 99WO-US12910.
 XX
 XX 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX
 XX (UYEM-) UNIV EMORY.
 PA
 XX Doetsch PW, Kaur B, Avery AM;
 PI
 XX WPI; 2000-1116417/10.
 DR N-PSDB; AAZ29860;
 XX
 XX A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers -
 PT
 XX Disclosure; Page 56; 133pp; English.
 XX
 CC The present sequence is a fusion protein comprising the GST signal
 CC peptide and delta228 truncated ultraviolet damage endonuclease (UVDE)
 CC from S. pombe. Delta228-UVDE is a 288 amino acid deletion of the
 CC N-terminal of the S. pombe uvel+ gene product. This is expressed in frame
 CC with a GST leader sequence. Stable endonuclease fragments can be produced
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, platinum diadduct, an intercalated
 CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 SQ Sequence 600 AA;

Query Match 100.0%; Score 1979; DB 21; Length 600;
 Best Local Similarity 100.0%; Pred. No. 4.8e-187;
 Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DDHAPREMFDCLDKPIPWGRGLGYACLTILSRMKERVFCSTCRITTIQDGLSVKQL 60
 Db 230 ddhapremfdcldkpwwgrgrlgyacintilrsmkervfcsrtcrittiqdglsvkql 289
 QY 61 GTQNVLDLILKLVENHNFGIHFMRVSSDLFPFASHAKYGYTLFQAQSHLEEVGKLANKYN 120
 Db 290 gtqnvldlilkvewnhnfgihfmrsvssdlfpfashakgytlfqaqshleevgklankyn 349
 QY 121 HRLTMHPGQYTOIASPREVVVDSATRDLAYHDEILSRMKLNQKNDVLIHILGGTFEG 180
 Db 350 hrltmhpggytqiasprevvvdsairdlayhdeilsrmlneqinkndavliihlggtfeg 409
 QY 181 KKETLDRFKNTQRLSDSVKARLVLENDVSVSDLLPLCOELNPLVLDWHHNIVPG 240
 Db 410 kketldrfknyqrlsdsvkarlvlendvsvsqdllplcqelnplvldwhhnivpg 469
 QY 241 TLREGSLDMLPIPTIRETWKGTQKOHYSESADPTAISGMKRRASDRVDFPFCDP 300
 Db 470 tlregslldmpliptiretwrkgtqkqhyssadptaisgmkrasdrvfdfppcdp 529
 QY 301 TMDLMEAKEKEQAVFELCRRYELONPCPLEIMGPEYDQTRDGYPPGAEKRLTARKRR 360
 Db 530 tmdlmeakekeqavfelcrryelnppcpleimgpeydtqrdgyppgaekrltarkrr 589
 QY 361 SRKEEVEDEK 371
 XX

Db 590 srkeeveedek 600

RESULT 3
 AAY44498
 ID AAY44498 standard; Protein; 828 AA.
 XX
 AC AAY44498;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Gsr signal peptide and S. pombe UVDE fusion protein.
 XX
 KW Gsr signal peptide; ultraviolet damage endonuclease; UVDE; Uvelp;
 KW UV irradiation; DNA damage; UV radiation damage; fusion protein;
 KW skin cancer; glutathione-S-transferase.
 XX
 OS Schizosaccharomyces pombe.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..229
 FT /label= GST_signal_peptide
 FT Protein 230..828
 FT /note= "S. pombe UVDE"
 FT Misc-difference 11
 FT /note= "Encoded by aaaa"
 FT Misc-difference 85
 FT /note= "Encoded by gtt"
 FT Misc-difference 147
 FT /note= "Encoded by gtt"
 FT
 XX
 PN WO9963828-A1.
 XX
 PD 16-DEC-1999.
 XX
 XX 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX
 PA (UYEM-) UNIV EMORY.
 XX
 XX Doetsch PW, Kaur B, Avery AM;
 XX
 XX WPI; 2000-116417/10.
 DR N-PSDB; AAZ29858.
 XX
 XX A new truncated ultraviolet damage endonuclease for treatment of skin
 XX cancers -
 XX
 PS Claim 7; Page 51; 133pp; English.
 XX
 CC The present sequence is a fusion protein comprising the Gsr signal
 CC peptide and full length ultraviolet damage endonuclease (UVDE) from
 CC Schizosaccharomyces pombe. UVDE genes are isolated by introducing a
 CC foreign cDNA library into a repair-deficient E. coli strain and
 CC selecting for complemented cells by UV irradiation of the transformants.
 CC This provides stable endonuclease fragments for cleaving a double-
 CC stranded DNA molecule that has a distorted structure resulting from UV
 CC radiation damage, a photoproduct, an abasic site, mismatched nucleotide
 CC pairing, platinum diadduct, an intercalated molecule or alkylation of a
 CC nucleotide. Uvelp can be used in compositions for internal or topical
 CC application and as a therapeutic agent for skin cancers.
 XX
 SQ Sequence 828 AA;

Query Match 100.0%; Score 1979; DB 21; Length 828;
 Best Local Similarity 100.0%; Pred. No. 7.9e-187;
 Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DDHAPREMFCDLKDPIPWGRGLGYACLATILRSMKERVFCSTCRITTIQDGLGVKQL 60

Db 458 ddhapremfcdldkdpipwgrglgyacntilrsmkervfcstcrittiqdglgvskql 517
 QY 61 GTQNVLDLKLVEVNHNFQIHFMVSSDLPFPASHAKYGYTLEFPAOSHLEEVGKLANKYN 120
 Db 518 gtqnvldliklvevnhnfgihfmrvssdlfpashakyygtlefaqshleevgkllankyn 577
 QY 121 HRLTHHPGOYTOIASPREVVVDSAIRDLAYHDEILSRMKLNQNLKNDKDAVLIIHLGCTFEG 180
 Db 578 hrlthhpqygtqlasprevvdsairdlayhdeilsrmklnqnlkndkdvliihlgctfeg 637
 QY 181 KKETLDRFRKNYQRLSDSVKARLVLENDVSVSDLLPLCQELNIPVLVDWHHNNIVPG 240
 Db 638 kketldrfknygrlsdsvkarlvlendvsvsqdlplcqelnipvlvdwhhnnivpg 697
 QY 241 TLREGSLDMLPLPTTRETWTTRKGIQKOHYSESADPTAISCMKRAHSDRVDFPPCDP 300
 Db 698 tlregslldmplpttretwttrkgitqkhyseadptaisgmkrhahsdrvdfppcdp 757
 QY 301 TMDLMIEAKEKEQAVFELCRRYELQNPCCPLEIMGPEDQTRDGYYPGCAEKRLTARKRR 360
 Db 758 tmdlmieakekeqavfelcrryelqnpccpleimgpeyqtrdgyypgaaekrlrtarkrr 817
 QY 361 SRKEEVEEDEK 371
 Db 818 srkeeveedek 828

RESULT 4
 AAW01618
 ID AAW01618 standard; Protein; 656 AA.
 XX
 AC AAW01618;
 XX
 DT 20-AUG-1997 (first entry)
 XX
 DE Neurospora crassa DNA repair enzyme.
 XX
 KW DNA repair enzyme; cyclobutane-type dimer; (6-4)-bound product;
 KW exposure; UV light; ultraviolet; suntan cream; prevention; treatment;
 KW skin cancer.
 XX
 OS Neurospora crassa.
 XX
 PN JP08266275-A.
 XX
 PD 15-OCT-1996.
 XX
 PF 29-MAR-1995; 95JP-0094137.
 XX
 PR 29-MAR-1995; 95JP-0094137.
 XX
 PA (SHIS) SHISEIDO CO LTD.
 XX
 DR WPI; 1996-512663/51.
 XX
 DR N-PSDB; AAT58286.
 XX
 PT DNA repair enzyme and related DNA - specifically recognises
 PT cyclobutane-type dimer and (6-4)-bound product induced by exposure
 PT to UV light
 XX
 PS Disclosure; Page 10-13; 18pp; Japanese.
 XX
 CC The present sequence shows a DNA repair enzyme which specifically
 CC recognises a cyclobutane-type dimer and (6-4)-bound product induced,
 CC respectively from TT and TC sequences in DNA, by exposure to UV light.
 CC The enzyme cleaves these sequences at the 5'-side. The DNA repair enzyme
 CC can be used in, e.g. suntan cream for prevention and treatment of skin
 CC cancer. Conventional endonuclease acts on either the cyclobutane-type
 CC dimer or the (6-4)-bound product in damaged DNA, while the enzyme
 CC of the invention can act on both of them.
 XX
 SQ Sequence 656 AA;

Query Match 48.5%; Score 960; DB 17; Length 656;
 Best Local Similarity 45.4%; Pred. No. 4.4e-86;
 Matches 205; Conservative 54; Mismatches 95; Indels 98; Gaps 10;

QY 15 PTPWGRGLGACLTNTILSMKERVFCSTCRITTI----- 49
 Db 201 plpwkgrlgayacintylrnskpffsrtcrmasivdhrhplqfedephehlnkpkdsk 260

QY 50 ---QRDGLSVKOLGTQNVLDLILKLVENHNFGIHPMRVSSDLFPFASHAKYGYTL-EFA 105
 Db 261 epqdelghkfvqelglanardivkmlew-----fpfashpvhgkylapfa 320

QY 106 QSHLEEVGLANKYNHRLTMHFGQYTOIASPREVVVDSVKARLVLENDVSWSDLLPLCQELN 165
 Db 321 sevlaeagrvaaelghritthpgqftqlgsprkevvesairdleyhdeallslkpeqgn 380

QY 166 KDAVLIHLGGTFEGCKKETLDLRFKKNYORLSDSVKARLVLENDVSWSDLLPLCQELN 225
 Db 381 rdavmlihmaggfgdkaatlerrfknyarlsqsknrlvlendvgwtvhdllpvcceeln 440

QY 226 IPLVLDWHHHNIV--PGLTREGSLDLM--PLIPTIRETWRTKGITQKHYESADPTAIS 281
 Db 441 lpmvldyhhnncfapahregtlidispkqieriantwkrkgikqkmhysepcd-gavt 499

QY 282 GMRRAHSDRVDFPPCDPTMDLMTAEKEQAVFELCRRYEL-----Q 325
 Db 500 prhrkrhrprvmtlppcpdmldmieakdkeqavfelmrftkllpgfekindmvpdydrdde 559

QY 326 NPPCF-----LEIMGPE-----YDQTRDG----- 344
 Db 560 nrpappvkapkkgkgrkrttdeaaapeevdtaaddvkdapegpkvpeeeramgpy 619

QY 345 ---YPPGAERKLTAARKRSRK-----EEVEED 369
 Db 620 nrvywplgceewlppkrevkkgvpeevede 651

RESULT 5
 AAY44502
 ID AAY44502 standard; Protein: 626 AA.
 AC AAY44502;
 XX
 DT 27-MAR-2000 (first entry)
 DE N. crassa delta228-UV damage endonuclease.
 KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
 KW glutathione-S-transferase signal peptide; uvel+ gene product;
 KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
 KW abasic site; platinum diaduct; mismatched nucleotide pairing;
 KW nucleotide alkylation; skin cancer.
 XX
 OS Neurospora crassa.
 XX
 PN WO9963828-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX
 PA (UYEM-) UNIV EMORY.
 XX
 PI Doetsch PW, Kaur B, Avery AM;
 XX
 DR WPI; 2000-116417/10.
 XX
 PT A new truncated ultraviolet damage endonuclease for treatment of skin

PT cancers -
 XX
 PS Claim 16; Page 59; 133pp; English.
 XX
 CC The present sequence is delta228-UV damage endonuclease from N. crassa.
 CC Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
 CC uvel+ gene product. This is expressed in frame with a GST leader sequence
 CC to generate a fusion protein. This provides stable endonuclease fragments
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, platinum diaduct, an intercalated
 CC molecule or alkylation of a nucleotide. uvelp can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 SQ Sequence 626 AA;

Query Match 46.0%; Score 910; DB 21; Length 626;
 Best Local Similarity 45.1%; Pred. No. 3.7e-81;
 Matches 197; Conservative 51; Mismatches 91; Indels 98; Gaps 10;

QY 15 PTPWGRGLGACLTNTILSMKERVFCSTCRITTI----- 49
 Db 201 plpwkgrlgayacintylrnskpffsrtcrmasivdhrhplqfedephehlnkpkdsk 260

QY 50 ---QRDGLSVKOLGTQNVLDLILKLVENHNFGIHPMRVSSDLFPFASHAKYGYTL-EFA 105
 Db 261 epqdelghkfvqelglanardivkmlew-----fpfashpvhgkylapfa 305

QY 106 QSHLEEVGLANKYNHRLTMHFGQYTOIASPREVVVDSVKARLVLENDVSWSDLLPLCQELN 165
 Db 306 sevlaeagrvaaelghritthpgqftqlgsprkevvesairdleyhdeallslkpeqgn 365

QY 166 KDAVLIHLGGTFEGCKKETLDLRFKKNYORLSDSVKARLVLENDVSWSDLLPLCQELN 225
 Db 366 rdavmlihmaggfgdkaatlerrfknyarlsqsknrlvlendvgwtvhdllpvcceeln 425

QY 226 IPLVLDWHHHNIV--PGLTREGSLDLM--PLIPTIRETWRTKGITQKHYESADPTAIS 281
 Db 426 lpmvldyhhnncfapahregtlidispkqieriantwkrkgikqkmhysepcd-gavt 484

QY 282 GMRRAHSDRVDFPPCDPTMDLMTAEKEQAVFELCRRYEL-----Q 325
 Db 485 prdrkrhrprvmtlppcpdmldmieakdkeqavfelmrftkllpgfekindmvpdydrdde 544

QY 326 NPPCF-----LEIMGPE-----YDQTRDG-----YPPGAERKLTA 356
 Db 545 nrpappvkapkkgkgrkrttdeaaapeevpeeeramgpynrvywplgceewlpp 604

QY 357 RKRRSRK-----EEVEED 369
 Db 605 kkrvkkkgvpeevede 621

RESULT 6
 AAY44503
 ID AAY44503 standard; Protein: 626 AA.
 XX
 AC AAY44503;
 XX
 DT 27-MAR-2000 (first entry)
 DE B. subtilis delta228-UV damage endonuclease.
 XX
 KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
 KW glutathione-S-transferase signal peptide; uvel+ gene product;
 KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
 KW abasic site; platinum diaduct; mismatched nucleotide pairing;
 KW nucleotide alkylation; skin cancer.
 XX
 OS Bacillus subtilis.
 XX

PN W09963828-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX
 PA (UYEM-) UNIV EMORY.
 XX
 PI Doetsch PW, Kaur B, Avery AM;
 XX
 DR WPI: 2000-116417/10.
 XX
 OS A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers -
 PT
 PS
 XX
 PF Claim 16; Page 59; 133pp; English.
 XX
 CC The present sequence is delta228-UV damage endonuclease from B. subtilis.
 CC Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
 CC uvel+ gene product. This is expressed in frame with a GST leader sequence
 CC to generate a fusion protein. This provides stable endonuclease fragments
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, a platinum diadduct, an intercalated
 CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 SQ Sequence 626 AA;

Query Match 46.0%; Score 910; DB 21; Length 626;
 Best Local Similarity 45.1%; Pred. No. 3.7e-81;
 Matches 197; Conservative 51; Mismatches 91; Indels 98; Gaps 10;
 QY 15 PIPNRGLGYACLTILRSMKERVFCSTCRITTI----- 49
 Db 201 plpwkgrlgyacintyrnakppifsrctmasivdhrhplqfedepehlnknpdksk 260
 QY 50 ---QRDGLSVKQLGTQNVLDLIKLVWNNHNFHFRVSSDLFPFASHAKYGYTL-EFA 105
 Db 261 epqdelghkfvgelglanardivkmclw-----fpfashpvhgkykiapfa 305
 QY 106 QSHLEEVGKLANKYNHRLTMHPGGYQTQIASPREVYVDSAIRDLAYHDEILSRMKLNQLN 165
 Db 306 sevlaeagrvaaelghrltthpgqftqgsprkevesairdleyhdelisliklpeqpn 365
 QY 166 KDAVLIILHGGTFEGKXETLDRFRKNYQRLSDSVKARLVLENDDVSVSDLLPLCQELN 225
 Db 366 rdavmihmg99qfgdkaatlerfkrnyarlsqsknrlvlenddvgvwtvhlipvceeln 425
 QY 226 IPLVDLHHHNNIV--PGLTREGSLDM--PLIPIITRETWTGRKGTOKOYHSESADPTAIS 281
 Db 426 ipmvlghhnnicfpahrlregtlidisdplqeriantwkrkikqkmhysepcd-gavt 484
 QY 282 GMKRRASDRVDFPPCPDPTMDLMEAKEQAVFELCRRYEL-----Q 325
 Db 485 prdrkrhrvmtlppcpdpmdlmieakkeqavfelmtfklpgfekindmvpardde 544
 QY 326 NPCC-----PLEIMGPEYDQTRDG-----YYPGGAERLTA 356
 Db 545 nrpappvkapkkkgkrkrttdeaaapeeveeramgpgynrvyplgceewlqp 604
 QY 357 RKRRSRK----EEVEED 369
 Db 605 kkrrevkkgkvpveede 621

RESULT 7
 AAY44505

ID AAY44505 standard; Protein; 294 AA.
 XX
 AC AAY44505;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE D. radiodurans (delta228-UV damage endonuclease).
 XX
 KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
 KW glutathione-S-transferase signal peptide; uvel+ gene product;
 KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
 KW abasic site; platinum diadduct; mismatched nucleotide pairing;
 KW nucleotide alkylation; skin cancer.
 XX
 OS Dienococcus radiodurans.
 XX
 PN W09963828-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX
 PA (UYEM-) UNIV EMORY.
 XX
 PI Doetsch PW, Kaur B, Avery AM;
 XX
 DR WPI: 2000-116417/10.
 XX
 OS A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers -
 PT
 PS
 XX
 PF Claim 16; Page 60; 133pp; English.
 XX
 CC The present sequence is delta228-UV damage endonuclease from the
 CC D. radiodurans. Delta228-UVDE is a 288 amino acid deletion of the
 CC N-terminal of the uvel+ gene product. This is expressed in frame with a
 CC GST leader sequence to generate a fusion protein. This provides stable
 CC endonuclease fragments for cleaving a double-stranded DNA molecule that
 CC has a distorted structure resulting from UV radiation damage, a
 CC photoproduct, an abasic site, mismatched nucleotide pairing, a platinum
 CC diadduct, an intercalated molecule or alkylation of a nucleotide. Uvelp
 CC can be used in compositions for internal or topical application and as a
 CC therapeutic agent for skin cancers.
 XX
 SQ Sequence 294 AA;

Query Match 11.6%; Score 230.5; DB 21; Length 294;
 Best Local Similarity 27.9%; Pred. No. 2.3e-14;
 Matches 88; Conservative 42; Mismatches 140; Indels 45; Gaps 10;
 QY 21 RLGYACLTILRSMKERVFCSTCRITTIORDGLSVKQLGTQNVLDLIKLVWNNHNFECI 80
 Db 1 qglvciltvgpevfrftvlsryalspaerea--kildlyssnktlrgaadycaahdi 58
 QY 81 HPMRVSSDLFP---FASHAKYGYTLEFAQSHLEEVGKLANKYNHRLTMHPGGYQTQIASPR 137
 Db 59 rlyrisslfpmlldlagddtgavvthlapqlleaghaftdagvrlmhpeqfivlnsdr 118
 QY 138 EVVVDISAIRDLAYHDEI-----LSRMKLNQNLKDAVLIILHGGTFEGKXETLDRFRKNY 192
 Db 119 pevressvramsaaharvmaglgartpwn-----lillhggkgrgae-----laali 166
 QY 193 QRLSDSVKARLVLENDDVSVSDLLPLCQELNIPVLVDWHHNNIVPGTLREGSLDLMF- 251
 Db 167 pdlipdvrlrlglenderayspaellpiceatgtplvfdaahhvv-----hdklpd 217
 QY 252 -LIPTTRETWTGRKGT-----QKOHYSESADPTAISGMKRRASDRVDFPPCPDPTM-D 303
 Db 218 qedpsvrewvlratwqppewgvvhls-----nglegpqdrtrhshlladftpsayadvpq 272

Qy 304 LMIEAKEKEQAVFEL 318
 Db 273 leveakgkeaaal 287

RESULT 8
 AAG41350
 ID AAG41350 standard; Protein: 388 AA.
 XX
 AC AAG41350;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 51435.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 18-MAY-1999; 99US-0134370.
 PR 19-MAY-1999; 99US-0134768.
 PR 20-MAY-1999; 99US-0134941.
 PR 21-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 23-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.

```

PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150366.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      5.5%  Score 108.5; DB 21; Length 388;
Best Local Similarity 20.0%; Pred. No. 0.042;
Matches 69; Conservative 60; Mismatches 123; Indels 93; Gaps 16;

QY 27 LNTILRSKKEVFE-----CSRTRCTTTTQDGLSVKQIGTGVNVLDTLKLVEWNH 76
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 54 mnsIargqkIplfsaaqlphneIaaqlcr-----qagI---vkrlI--ektvdIle---dh 100
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY 77 ---NFGIHFMRVSSDLFPFASHAKYGYTLEFAQSHLEEVG-----KLANK----- 118
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```

```

Db 101 gednfaivfaamgvnm-----etaqffkfrdfceengsmervtllflinlandptieri 150
QY 119 YNHRUTMHPGQYQTIASPREVV-----DSAIRDL-AYHDEILSRMKLNEQLNKDAV 169
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 151 itprIalttaeylayecgkhviltmdssyadali revsaareevpgrgypgmytdlia 210
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 170 LIIHLGGTFEaiaKFTLDRFRKNYQRLSDSVKARLVLENDVSVQDILLPLCOELNIPLV 229
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 211 tiyeragriegrksitqi-----pltmpnddithptpditgyitegqiylid 258
QY 230 LDWHHHNIVPGTLREGSLDMLPIPTIRETWTRKGITOKQH-----YSESADPTAISG 282
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 259 rqlhnrqiyp-----pinvlpslrmlksaigegmtrkdhdsvnqlyanyaigkdvqa 312
QY 283 MKRRAHSRDRVDFPPCDPTMDLM-IEAKEKEQAVFLCRRRYELQN 326
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 313 mkavvgeaal-----ssedllylefldkferkfvmqgaydlrn 350

RESULT 9
AAG41349
ID AAG41349 standard; Protein; 470 AA.
XX
AC AAG41349;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51434.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.

```

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139432.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142330.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.

[illegible]

PR 22-JUL-1999;	99US-0145089.	PR 21-OCT-1999;	99US-0160814.
PR 22-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145216.	PR 22-OCT-1999;	99US-0160981.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160989.
PR 26-JUL-1999;	99US-0145276.	PR 25-OCT-1999;	99US-0161404.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161406.
PR 27-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161359.
PR 28-JUL-1999;	99US-0145931.	PR 26-OCT-1999;	99US-0161360.
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146388.	PR 28-OCT-1999;	99US-0161920.
PR 02-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161992.
PR 03-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161993.
PR 04-AUG-1999;	99US-0147204.	PR 29-OCT-1999;	99US-0162142.
PR 04-AUG-1999;	99US-0147302.		
PR 05-AUG-1999;	99US-0147320.		
PR 05-AUG-1999;	99US-0147260.		
PR 06-AUG-1999;	99US-0147303.		
PR 06-AUG-1999;	99US-0147416.		
PR 09-AUG-1999;	99US-0147493.		
PR 09-AUG-1999;	99US-0147935.		
PR 10-AUG-1999;	99US-0148171.		
PR 11-AUG-1999;	99US-0148319.		
PR 12-AUG-1999;	99US-0148341.		
PR 13-AUG-1999;	99US-0148565.		
PR 13-AUG-1999;	99US-0148684.		
PR 16-AUG-1999;	99US-0149368.		
PR 17-AUG-1999;	99US-0149175.		
PR 18-AUG-1999;	99US-0149426.		
PR 20-AUG-1999;	99US-0149722.		
PR 20-AUG-1999;	99US-0149723.		
PR 20-AUG-1999;	99US-0149929.		
PR 23-AUG-1999;	99US-0149902.		
PR 23-AUG-1999;	99US-0149930.		
PR 25-AUG-1999;	99US-0150566.		
PR 26-AUG-1999;	99US-0150884.		
PR 27-AUG-1999;	99US-0151065.		
PR 27-AUG-1999;	99US-0151066.		
PR 27-AUG-1999;	99US-0151080.		
PR 30-AUG-1999;	99US-0151303.		
PR 31-AUG-1999;	99US-0151438.		
PR 01-SEP-1999;	99US-0151930.		
PR 07-SEP-1999;	99US-0152363.		
PR 10-SEP-1999;	99US-0153070.		
PR 13-SEP-1999;	99US-0153758.		
PR 15-SEP-1999;	99US-0154018.		
PR 16-SEP-1999;	99US-0154039.		
PR 20-SEP-1999;	99US-0154779.		
PR 22-SEP-1999;	99US-0155139.		
PR 23-SEP-1999;	99US-0155486.		
PR 24-SEP-1999;	99US-0155659.		
PR 28-SEP-1999;	99US-0156458.		
PR 29-SEP-1999;	99US-0156596.		
PR 04-OCT-1999;	99US-0157117.		
PR 05-OCT-1999;	99US-0157755.		
PR 06-OCT-1999;	99US-0157865.		
PR 07-OCT-1999;	99US-0158029.		
PR 08-OCT-1999;	99US-0158232.		
PR 12-OCT-1999;	99US-0158369.		
PR 13-OCT-1999;	99US-0159293.		
PR 13-OCT-1999;	99US-0159294.		
PR 13-OCT-1999;	99US-0159295.		
PR 14-OCT-1999;	99US-0159329.		
PR 14-OCT-1999;	99US-0159330.		
PR 14-OCT-1999;	99US-0159331.		
PR 14-OCT-1999;	99US-0159637.		
PR 14-OCT-1999;	99US-0159638.		
PR 18-OCT-1999;	99US-0159584.		
PR 21-OCT-1999;	99US-0160741.		
PR 21-OCT-1999;	99US-0160767.		
PR 21-OCT-1999;	99US-0160768.		
PR 21-OCT-1999;	99US-0160770.		
PR 22-OCT-1999;			

Query Match 5.5%; Score 108.5; DB 21; Length 486;

Best Local Similarity 20.0%; Pred. No. 0.059; Mismatches 60; Conservative 69; Indels 93; Gaps 16;

Matches	69	Conservative	60	Mismatches	123	Indels	93	Gaps	16;
QY	27	LNTILSRKERVF-----	CSRTCTRTTQRDGLSVKQLGTONVLDLIKLVENH	76					
Db	152	msiargqkipifsaaglpheiaaqicr-----	gagl--vkrl--ektvdlle----	dh	198				
QY	77	---NFGIHPMRVSSDLFPFASHAKYGYTLEFAQSHLEEVG-----	KLANK-----	118					
Db	199	gednfaivfaamgvnm-----	etaqfkkrdfceengsmervtlfInlandptieri	248					
QY	119	YNHRLTMHFGQYQIASPREVVV-----	DSAIRDL--AYHDEILSRMKLNQOLNKDAV	169					
Db	249	itprialtaeylayecgkhviltmssyadairvsaareevpgrirgypgmyvldla	308						
QY	170	LIHLGGTEGKKETLDRFRKKNYQRLSDSVKARLVLEDDVSVQDQLPLCQELNIPLV	229						
Db	309	tiyeragriegrkgsitql-----	pillmpnddithptdltgyiteggyid	356					
QY	230	LDWHHNIVPGTLREGSLDMLPLIPTIRETWTRKGTQKQH-----	YESADPTAISG	282					
Db	357	rqlhnrqlyp-----	pinvlpslsrlmksaigemtrkhdsvsnqlyanyalgkdvqa	410					
QY	283	MKRRASDRVDFPPCDPTMDLM--IEAKEKEQAVFELCRRYELQN	326						
Db	411	mkavvqeal-----	ssedllylefldkferkfvmqaydtrn	448					

RESULT 11

AAG18697

ID AAG18697 standard; Protein; 615 AA.

XX AAG18697;

AC AAG18697;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 20209.

DE Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

PN HP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138340.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.

```

PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162143.

Query Match      5.4%; Score 107.5; DB 21; Length 615;
Best Local Similarity 22.6%; Pred. No. 0.1;
Matches 61; Conservative 34; Mismatches 98; Indels 77; Gaps 11;

Qy 48 TIORCLESVKOL-----GTONVLDLKLVEWHNFCIHPMYSSDLFPASHAKYCYT 101
Db 344 timeqdlrldqleqdvfgdagrkdvikflstnh-----iish----- 362

Qy 102 LEFAASHLEEVGKLANKNYHRLTMHPGQYQIASPREVVVDSAIRDLAYHDEILSRMKN 161
Db 383 ----eskl-----NIPVLVDWHHNIIVPGTLREGSL--DLMLPLIPIRE--TWTRKGI 265

Qy 162 EQLNKDAVLIIH---LGTF-EGKKTLDLRFKKNYQELSDSVKARLVLENDVSNVQD 216
Db 412 aksldgddvavnmrlilgpyhteckstgtsfpikfdvktkraardrvgetqqlsr 471

Qy 217 LPLCQEL-----NIPVLVDWHHNIIVPGTLREGSL--DLMLPLIPIRE--TWTRKGI 265
Db 472 fypiveelvelkiskghlp-kdypcmnepkptfysgslspsasvplphsrrtptwarhl 530

Qy 266 TOKQHYSESADPTAISGMKRAHSDRVDF 295
Db 531 sddgyfsdsvlgrassgfkfk-ggrifvf 558

RESULT 12
AAG50810
ID AAG50810 standard; Protein; 615 AA.
XX
AC AAG50810;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 64427.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.

```

```

XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.

```



```
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
```

```
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.4%; Score 107.5; DB 21; Length 615;
Best Local Similarity 22.6%; Pred. No. 0.1;
Matches 61; Conservative 34; Mismatches 98; Indels 77; Gaps 11;

QY 48 TIORDGLESVKOL-----GTQNVLDLKLVEWNNHFGIHFRVSSDLFPFASHAKYGYT 101
Db 344 timeqglrdlqleqdlvfgdagrkdvkflstnh-----ilish----- 382
QY 102 LEFAQSHLEEVGKLANKYNHRLTMHPGOYTOIASPREVVVDSAIRDLAYHDEILSRMK1N 161
Db 383 ----eskl-----rllimivaaipkpkfegekgrkmmel 411
QY 162 EQLNKDAVLIIH----LGQTF--EGKKETLDRFRKNYQRLSDSVKARLVLENDVSNVOD 216
Db 412 aklsqddvvavnmrllgpvhlecksttgsfpkfdvltkrraarrrdvgetqtwlsr 471
QY 217 LLPLQCEL-----NIPLVLDWHHHNIVPGTLREGSL--DLMPILPTIRE--TWTRKGI 265
Db 472 fypiveelweklsknhlp--kqyccmepkpfysgsispsasplphsrrtptwarhl 530
QY 266 TQKHYESADPTAISGMKRRASDRVDFD 295
Db 531 sddgyfsdsvlgrassgfrk--gqrlvfv 558
```

RESULT 13

AAG18696

ID AAG18696 standard; Protein; 633 AA.

XX AAG18696;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 20208.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

OS Arabidopsis thaliana.

XX EPI033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139452.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

99US-0140354.

99US-0140695.

99US-0140823.

99US-0140991.

99US-0141287.

99US-0141842.

99US-0142154.

99US-0142055.

99US-0142390.

99US-0142803.

99US-0142920.

99US-0142977.

99US-0143542.

99US-0143624.

99US-0144005.

99US-0144085.

99US-0144086.

99US-0144325.

99US-0144331.

99US-0144332.

99US-0144333.

99US-0144334.

99US-0144335.

99US-0144352.

99US-0144632.

99US-0144884.

99US-0144884.

99US-0145086.

99US-0145088.

99US-0145085.

99US-0145087.

99US-0145089.

99US-0145192.

99US-0145145.

99US-0145218.

99US-0145224.

99US-0145276.

99US-0145913.

99US-0145918.

99US-0145919.

99US-0145951.

99US-0146386.

99US-0146388.

99US-0146389.

99US-0147038.

99US-0147204.

99US-0147302.

99US-0147192.

99US-0147260.

99US-0147303.

99US-0147416.

99US-0147493.

99US-0147935.

99US-0147935.

99US-0148171.

99US-0148319.

99US-0148341.

99US-0148565.

99US-0148684.

99US-0149368.

99US-0149175.

99US-0149426.

99US-0149722.

99US-0149723.

99US-0149929.

99US-0149902.

99US-0149930.

99US-0150566.

99US-0150884.

99US-0151065.

99US-0151066.

99US-0151080.

99US-0151303.

99US-0151348.

PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	18-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	21-JUN-1999;	99US-0139817.	PR	30-AUG-1999;	99US-0151303.
PR	22-JUN-1999;	99US-0139899.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140353.	PR	01-SEP-1999;	99US-0151930.
PR	23-JUN-1999;	99US-0140354.	PR	07-SEP-1999;	99US-0152363.
PR	24-JUN-1999;	99US-0140695.	PR	10-SEP-1999;	99US-0153070.
PR	28-JUN-1999;	99US-0140823.	PR	13-SEP-1999;	99US-0153758.
PR	29-JUN-1999;	99US-0140991.	PR	15-SEP-1999;	99US-0154018.
PR	30-JUN-1999;	99US-0141287.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0141842.	PR	20-SEP-1999;	99US-0154779.
PR	01-JUL-1999;	99US-0142154.	PR	22-SEP-1999;	99US-0155139.
PR	02-JUL-1999;	99US-0142053.	PR	23-SEP-1999;	99US-0155486.
PR	06-JUL-1999;	99US-0142390.	PR	24-SEP-1999;	99US-0155659.
PR	08-JUL-1999;	99US-0142803.	PR	28-SEP-1999;	99US-0156458.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156596.
PR	12-JUL-1999;	99US-0142977.	PR	04-OCT-1999;	99US-0157117.
PR	13-JUL-1999;	99US-0143542.	PR	05-OCT-1999;	99US-0157753.
PR	14-JUL-1999;	99US-0143624.	PR	06-OCT-1999;	99US-0157865.
PR	15-JUL-1999;	99US-0144005.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	16-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144325.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144333.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159330.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159638.
PR	20-JUL-1999;	99US-0144884.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0144814.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160767.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160815.
PR	22-JUL-1999;	99US-0145192.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160989.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0161040.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145918.	PR	26-OCT-1999;	99US-0161359.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161360.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146386.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161992.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-01

PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 5.4%; Score 107.5; DB 21; Length 662;
 Best Local Similarity 22.6%; Pred. No. 0.12;
 Matches 61; Conservative 34; Mismatches 98; Indels 77; Gaps 11;
 QY 48 TIGDGLSVKOL-----GTQNVLDLKLVEWNNHFGIHFMRVSSDLFPFASHAKYGT 101
 Db 391 timeqglrdlqleqdlvfgdagrkdvikfisthh-----lsh----- 429

QY 102 LEPAQSHLEEVGKLANKNHRLTMHFGQYQIASPREVVVDSAIRDLAYHDEILSRMKLN 161
 Db 430 -----eskl-----rlmivaaiypkkfegekgrkmmel 458
 QY 162 EQLNKDAVLIH-----LGSTF-EGK*ETLD*FRKKNYQR*LS*SVKAR*LVLENDDV*SV*OD 216
 Db 459 akslgddvavnnmrlligpvhtecksttgsfplkfdvltkrraarrdrvgetqtwqjsr 518
 QY 217 LLPLCQEL-----NIPLVLDWHHHNIVPGTLREGSL--DLMPLIPTIRE--TWTRKGI 265
 Db 519 fypiveelveklskghlp-kqdyppcmnekpftfysqslspsaspvlphsrrtptwarthl 577
 QY 266 TOKQHYSESADPTAISGMKRRRAHS*DRV*DF 295
 Db 578 sddgyfsdsvlgrassgfkrrk--ggrifvf 605

Search completed: January 15, 2002, 13:50:00
 Job time: 289 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:50:15 ; Search time 120.01 Seconds
(without alignments)
181.464 Million cell updates/sec

Title: US-09-724-296-39

Perfect score: 1527

Sequence: 1 QLGVLCTVGPVRFRTVT.....VEAKGKEAIAALRLMAPFK 294

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1527	100.0	294	21	AA1980
2	255.5	16.7	656	17	AA1981
3	239.5	15.7	626	21	AA1982
4	239.5	15.7	626	21	AA1983
5	230.5	15.1	371	21	AA1984
6	230.5	15.1	600	21	AA1985
7	230.5	15.1	828	21	AA1986
8	94.5	6.2	270	14	AA1987
9	93	6.1	199	19	AA1988
10	92.5	6.1	915	22	AA1989
11	90	5.9	497	22	AA1990

12	90	5.9	497	22	AA1991
13	90	5.9	1421	21	AA1992
14	90	5.9	1891	19	AA1993
15	90	5.9	3413	19	AA1994
16	89	5.8	6797	22	AA1995
17	88	5.8	537	18	AA1996
18	86	5.6	1110	22	AA1997
19	85.5	5.6	653	17	AA1998
20	84.5	5.5	613	22	AA1999
21	84.5	5.5	616	22	AA2000
22	84.5	5.5	646	22	AA2001
23	84	5.5	403	20	AA2002
24	83	5.4	1291	20	AA2003
25	83	5.4	1891	18	AA2004
26	83	5.4	1391	18	AA2005
27	82.5	5.4	1160	21	AA2006
28	82.5	5.4	1160	22	AA2007
29	82.5	5.4	1160	22	AA2008
30	82.5	5.4	3567	14	AA2009
31	82	5.4	404	10	AA2010
32	82	5.4	1477	20	AA2011
33	82	5.4	1494	20	AA2012
34	81.5	5.3	198	22	AA2013
35	81.5	5.3	465	22	AA2014
36	81.5	5.3	5588	20	AA2015
37	81.5	5.3	5588	22	AA2016
38	81	5.3	506	9	AA2017
39	81	5.3	521	9	AA2018
40	81	5.3	607	18	AA2019
41	80.5	5.3	186	21	AA2020
42	80.5	5.3	187	21	AA2021
43	80.5	5.3	221	21	AA2022
44	80.5	5.3	221	21	AA2023
45	80.5	5.3	249	21	AA2024

ALIGNMENTS

RESULT 1

AA1980

ID AA1980 standard; Protein: 294 AA.

AC AA1980

XX AA1980

DT 27-MAR-2000 (first entry)

XX

XX

DE D. radiodurans delta228-UV damage endonuclease.

XX Delta228-UVDE: ultraviolet damage endonuclease; GST signal peptide;

KW glutathione-S-transferase signal peptide; uvex-gene product;

KW UV irradiation; DNA damage; UV radiation damage; photoproduct;

KW abasic site; apurinic diaduct; mismatched nucleotide pairing;

KW nucleotide alkylation; skin cancer.

XX

OS *Dienococcus radiodurans*.

XX

XX WO9953828-A1.

XX

XX 16-DEC-1999.

XX

PF 08-JUN-1999; 99WO-US12910.

XX

PR 08-JUN-1998; 98US-008521.

PR 18-MAY-1999; 99US-0134752.

XX

PA (UYEM-) UNIV EMORY.

XX

PI Doetsch PW, Kaur B, Avery AM;

XX WPI; 2000-116417/10.

XX

PT A new truncated ultraviolet damage endonuclease for treatment of skin

PT cancers -
 XX Claim 16; Page 60; 133pp; English.
 PS
 XX The present sequence is delta228-UV damage endonuclease from
 CC D. radiodurans. Delta228-UVDE is a 288 amino acid deletion of the
 CC N-terminal of the uvel+ gene product. This is expressed in frame with a
 CC GST leader sequence to generate a fusion protein. This provides stable
 CC endonuclease fragments for cleaving a double-stranded DNA molecule that
 CC has a distorted structure resulting from UV radiation damage, a
 CC photoproduct, an abasic site, mismatched nucleotide pairing, a platinum
 CC diadduct, an intercalated molecule or alkylation of a nucleotide. Uvelp
 CC can be used in compositions for internal or topical application and as a
 CC therapeutic agent for skin cancers.
 XX
 SQ Sequence 294 AA;

Query Match 100.0%; Score 1527; DB 21; Length 294;
 Best Local Similarity 100.0%; Pred. No. 3e-159;
 Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLCVCLTVGPEVRFRTVLSRYALSPAEREAKLLDLYSSNIKTLRGAADYCAAHDIRL 60
 DB 1 qglvcltvgepvrfrtvlrsyalspaereaklldlyssniktlrqaadycaahdirl 60

QY 61 YRLSSSLFPMLDLAGDDTCAAVLTHLAPOLLEAGHAFTDAGVRLLMHPPEQFTVLNSDRPE 120
 DB 61 yrlssslfpmldlagdgtcaavtlhapqlleaghaftdagvrlmhppeqftvlnsdrpe 120

QY 121 VRESSVRAMSAHARVMDGLGARTPNWLLHLLHGGKGGAGAEALALIPDLPPVRLRLGLE 180
 DB 121 vressvrmsaharvmdgllgartpnwllllhggkggagaelaalipdlppvrlrlgle 180

QY 181 NDERAYSPAELLIPCEATGTPLVFDAAHHVVDKLPDQEDPSVREWVLRARATWOPPEWQ 240
 DB 181 nderayspaellipceatgtplvfdahhhvvdhklpqdqedpsvrevvlaratwqppewq 240

QY 241 VVHLSNGIEGPDQRHSHLIAFPAYADVPOIEVEAKGKEEAIAALRLMAPFK 294
 DB 241 vvhlslngiegpqdrhshliadfpayadvpqieveakgkeeaiaalrlmapfk 294

RESULT 2
 AAW01618
 ID AAW01618 standard; Protein; 656 AA.
 XX
 AC AAW01618;
 XX
 DT 20-AUG-1997 (first entry)
 XX
 DE Neurospora crassa DNA repair enzyme.
 XX
 KW DNA repair enzyme; cyclobutane-type dimer; (6-4)-bound product;
 KW exposure; UV light; ultraviolet; suntan cream; prevention; treatment;
 KW skin cancer.
 XX
 OS Neurospora crassa.
 XX
 PN JP08266275-A.
 XX
 PD 15-OCT-1996.
 XX
 PE 29-MAR-1995; 95JP-0094137.
 XX
 PR 29-MAR-1995; 95JP-0094137.
 XX
 PA (SHIS) SHISEIDO CO LTD.
 XX
 DR WPI; 1996-512663/51.
 DR N-PSDB; AAT58286.
 XX
 PT DNA repair enzyme and related DNA - specifically recognises

PT cyclobutane-type dimer and (6-4)-bound product induced by exposure
 PT to UV light
 XX
 PS Disclosure; Page 10-13; 18pp; Japanese.
 XX
 CC The present sequence shows a DNA repair enzyme which specifically
 CC recognises a cyclobutane-type dimer and (6-4)-bound product induced,
 CC respectively from TT and TC sequences in DNA, by exposure to UV light.
 CC The enzyme cleaves these sequences at the 5'-side. The DNA repair enzyme
 CC can be used in, e.g. suntan cream for prevention and treatment of skin
 CC cancer. Conventional endonuclease acts on either the cyclobutane-type
 CC dimer or the (6-4)-bound product in damaged DNA, while the enzyme
 CC of the invention can act on both of them.
 XX
 SQ Sequence 656 AA;

Query Match 16.7%; Score 255.5; DB 17; Length 656;
 Best Local Similarity 31.8%; Pred. No. 5.3e-19;
 Matches 84; Conservative 38; Mismatches 103; Indels 39; Gaps 13;

QY 56 HDRLRYRLSSSLFPMLDLAGDDTCAAVLTHLAPOLL-BAGHAFTDAGVRLLMHPPEQFIVL 114
 DB 292 ygirflrlsseimp---fashpvhyklapfasevtaeagrvaaelghrltthpgqftql 348

QY 115 NSDRPEVRESSVRAMSAHARVMDGLGARTPNW---LLLHGGKGGAGAEALALIT--- 166
 DB 349 gspkvevsaiairdleyldel---lsllkpeqnrdaavmlih-mggqfgdkaatlerrf 403

QY 167 ---PDLPPVRLRLGLENDERAYSPAELLIPCEATGTPLVFDAAHHV---HDKL 215
 DB 404 knyarlsgqckarlvlendddvgwtvhdllpvcceelnipmvdlyhhnfcfdpahiregt 463

QY 216 PDQEDPSVREWVLRARATWOPPE-WQVHLSNGIEG---PQDRK-HSHLIAFPSPAYADV 270
 DB 464 ldsdpklqe---riantwkrkqikqkmhysepcdjavtprhrkrprvmtlppcpddm 520

QY 271 PQIEVEAKGKEEAIAALRLMAPFK 294
 DB 521 -dimieakdkeqav--felartfk 541

RESULT 3
 AAY44502
 ID AAY44502 standard; Protein; 626 AA.
 XX
 AC AAY44502;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE N. crassa delta228-UV damage endonuclease.
 XX
 KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
 KW glutathione-S-transferase signal peptide; uvel+ gene product;
 KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
 KW abasic site; platinum diadduct; mismatched nucleotide pairing;
 KW nucleotide alkylation; skin cancer.
 XX
 OS Neurospora crassa.
 XX
 PN WO9963828-A1.
 XX
 PD 16-DEC-1999.
 XX
 PE 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX
 PA (UYEW-) UNIV EMORY.
 XX
 PT Boetsch PW, Kaur B, Avery AM;

DR WPI; 2000-116417/10.
XX A new truncated ultraviolet damage endonuclease for treatment of skin
PT cancers -
PT
XX
XX Claim 16; Page 59; 133pp; English.
PS
XX
XX The present sequence is delta228-UV damage endonuclease from B. subtilis.
CC Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
CC uvel+ gene product. This is expressed in frame with a GST leader sequence
CC to generate a fusion protein. This provides stable endonuclease fragments
CC for cleaving a double-stranded DNA molecule that has a distorted
CC structure resulting from UV radiation damage, a photoproduct, an abasic
CC site, mismatched nucleotide pairing, a platinum diadduct, an intercalated
CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
CC for internal or topical application and as a therapeutic agent for skin
CC cancers.
XX
XX Sequence 626 AA;

Query Match 15.7%; Score 239.5; DB 21; Length 626;
Best Local Similarity 32.5%; Pred. No. 2.8e-17;
Matches 77; Conservative 34; Mismatches 87; Indels 39; Gaps 12;
QY 86 LAP-QLLEAGHAFTDAGVRLMHPEQFIVLNSDRPEVRESSVRAMSAHARVMDGLGL 141
DB 301 lapfasevlaeagrvaaelghrlthpghftqlgsprkeevvesairdleyhdel---lsl 357
QY 142 ARTPNW-----LLLLHGGKGGRAELAAII-----PDLDPVRLRLGLENDERAYS 189
DB 358 lkipeqndavmliih--mgggfgdkaatlfrknyarlsgscknrivlendvgwtvh 415
QY 190 ELLPICEATGTLVDFDAHHVV-----HDKLPQEDPSVREWVLRARATWQPP- 241
DB 416 dlipvecelnlpvldyhhnfcfpahrlregtldisdpkqkqk 472
QY 242 VHLNSGTEG---PQDRR-HSHLIADFPFAYADVPQIEVEAKGKEBAIAALRLMAPFK 294
DB 473 mhysepcdgavtprdrkrhrprvmtlppcpddm-dlmieakdkedav--feimrtfk 526

RESULT 5
AA44499
ID AAY44499 standard; Protein: 371 AA.
XX
XX AC AAY44499;
XX
XX DT 27-MAR-2000 (first entry)
XX
XX DE S. pombe delta228-UV damage endonuclease.
XX
XX KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
KW glutathione-S-transferase signal peptide; uvel+ gene product;
KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
KW abasic site; aplatium diadduct; mismatched nucleotide pairing;
KW nucleotide alkylation; skin cancer.
XX
XX OS Schizosaccharomyces pombe.
XX
XX PN WO9963828-A1.
XX
XX PD 16-DEC-1999.
XX
XX PF 08-JUN-1999; 99WO-US12910.
XX
XX PR 08-JUN-1998; 98US-0088521.
XX
XX PR 18-MAY-1999; 99US-0134752.
XX
XX PA (UYEM-) UNIV EMORY.
XX
XX PI Doetsch PW, Kaur B, Avery AM;
XX

DR WPI; 2000-116417/10.
XX A new truncated ultraviolet damage endonuclease for treatment of skin
PT cancers -
PT
XX
XX Claim 16; Page 59; 133pp; English.
PS
XX
XX The present sequence is delta228-UV damage endonuclease from N. crassa.
CC Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
CC uvel+ gene product. This is expressed in frame with a GST leader sequence
CC to generate a fusion protein. This provides stable endonuclease fragments
CC for cleaving a double-stranded DNA molecule that has a distorted
CC structure resulting from UV radiation damage, a photoproduct, an abasic
CC site, mismatched nucleotide pairing, a platinum diadduct, an intercalated
CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
CC for internal or topical application and as a therapeutic agent for skin
CC cancers.
XX
XX Sequence 626 AA;

Query Match 15.7%; Score 239.5; DB 21; Length 626;
Best Local Similarity 32.5%; Pred. No. 2.8e-17;
Matches 77; Conservative 34; Mismatches 87; Indels 39; Gaps 12;
QY 86 LAP-QLLEAGHAFTDAGVRLMHPEQFIVLNSDRPEVRESSVRAMSAHARVMDGLGL 141
DB 301 lapfasevlaeagrvaaelghrlthpghftqlgsprkeevvesairdleyhdel---lsl 357
QY 142 ARTPNW-----LLLLHGGKGGRAELAAII-----PDLDPVRLRLGLENDERAYS 189
DB 358 lkipeqndavmliih--mgggfgdkaatlfrknyarlsgscknrivlendvgwtvh 415
QY 190 ELLPICEATGTLVDFDAHHVV-----HDKLPQEDPSVREWVLRARATWQPP- 241
DB 416 dlipvecelnlpvldyhhnfcfpahrlregtldisdpkqkqk 472
QY 242 VHLNSGTEG---PQDRR-HSHLIADFPFAYADVPQIEVEAKGKEBAIAALRLMAPFK 294
DB 473 mhysepcdgavtprdrkrhrprvmtlppcpddm-dlmieakdkedav--feimrtfk 526

RESULT 4
AA44503
ID AAY44503 standard; Protein: 626 AA.
XX
XX AC AAY44503;
XX
XX DT 27-MAR-2000 (first entry)
XX
XX DE B. subtilis delta228-UV damage endonuclease.
XX
XX KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
KW glutathione-S-transferase signal peptide; uvel+ gene product;
KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
KW abasic site; aplatium diadduct; mismatched nucleotide pairing;
KW nucleotide alkylation; skin cancer.
XX
XX OS Bacillus subtilis.
XX
XX PN WO9963828-A1.
XX
XX PD 16-DEC-1999.
XX
XX PF 08-JUN-1999; 99WO-US12910.
XX
XX PR 08-JUN-1998; 98US-0088521.
XX
XX PR 18-MAY-1999; 99US-0134752.
XX
XX PA (UYEM-) UNIV EMORY.
XX
XX PI Doetsch PW, Kaur B, Avery AM;
XX

Query Match 6.28; Score 94.5; DB 14; Length 270;
Best Local Similarity 27.58; Pred. NO. 0.068;
Matches 70; Conservative 29; Mismatches 67; Indels 89; Gaps 17;

Tue Jan 15 14:12:19 2002

from Corynebacterium, and identifying a homologue of a gene derived from corynebacterium. Corynebacterium bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a protein described in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

XX SQ Sequence 497 AA;

Query Match 5.9%; Score 90; DB 22; Length 497;
Best Local Similarity 23.5%; Pred. No. 0.54;
Matches 72; Conservative 31; Mismatches 90; Indels 114; Gaps 14;

QY 20 LSKYRALSPAEREAKLLDLYS-----SNIKTLRGAADYCAAHDIRLY 61
DB 68 LKWSALTGAERTGYLKIATELESSEALALNTRENGSPISETRGAASNAAG----if 123
QY 62 RLSSSLFPMLDLAGDD-----TGAA-----VLTHLAPOLLE 92
DB 124 RYFATLAPWLD--GEDIRFPAGSAESIVDKDPGVCAIAPWNTFPIINVIKLAPALL- 180
QY 93 AGHAFTDAGVRLLMHPE-----QFIVLNSDRPEVRESSVRAMSAHARVMDGLGLARTP 145
DB 181 -----AGCTVVIKPSAPTPLSIRFILAELAAGVPGVNVLLTSGRFGD--alvrhp 231
QY 146 WNLHLHGKGGKGAELAAALPDLPPVRLRLGLENDERAYSPAELLPTCEATGTPLVFD 205
DB 232 GVDKVAFTGATPGVKKIAACGELLRPVTLIELG-----GKSSAILP-----d 274
QY 206 AHHVHVHDKLPDQEDPSVRWMLRARATWQPQWVHLNSGIEGPDREHSHLIADFFS 265
DB 275 admsvlstrl-----irscm-----rnt-----gqcyistrilapsr----- 308
QY 266 AYADVPQ 272
DB 309 -yaevvq 314

RESULT 12

AAB79113
ID AAB79113 standard; Protein; 497 AA.

XX AC AAB79113;

XX DT 30-APR-2001 (first entry)

XX DE Corynebacterium glutamicum HA protein sequence SEQ ID NO:182.

XX KW Corynebacterium glutamicum; homeostasis; adaptation; HA protein;
fine chemical production; organic acid; proteinogenic amino acid;
nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
diagnosis; Corynebacterium diphtheriae; genetic engineering;
Brevibacterium; environmental condition.

XX OS Corynebacterium glutamicum.

XX PN WO200100842-A2.

XX PD 04-JAN-2001.

XX PF 23-JUN-2000; 2000WO-1B00911.

XX PR 25-JUN-1999; 99US-0141031.

XX PR 08-JUL-1999; 99DE-1031636.

XX PR 09-JUL-1999; 99DE-1032125.

XX PR 09-JUL-1999; 99DE-1032126.

XX PR 09-JUL-1999; 99DE-1032127.

XX PR 09-JUL-1999; 99DE-1032128.

(PTDD) that induces apoptosis. p53 can hold cells in check, quiescent and prevent cells from becoming cancerous. PTDD comprises leucine-rich repeat (LRR), antigenic fragments and p53-binding sequence which is responsive to p53. PTDD is useful for treating or preventing diseases associated with cell cycle regulation, cell growth and deficient or insufficient apoptosis, autoimmune disease, neurodegenerative disorder, stroke, cancer, dysplasia and Li-Fraumeni syndrome.

XX SQ Sequence 915 AA;

Query Match 6.1%; Score 92.5; DB 22; Length 915;
Best Local Similarity 26.7%; Pred. No. 0.72;
Matches 43; Conservative 29; Mismatches 62; Indels 27; Gaps 9;

QY 80 AAVLTHLAPOLLEA-----GHAFDAGVRLLM-----HPFQFVLSNDRPEVRESSVR 127
DB 14 AAAAADAATLEAVDAGPGAPFPAQGNLIDLRPGGCHRLQYL-CSQPPQLIQVEFL 72
QY 128 AMSAH--ARYMDGLGIARTPWNL-----LLHGG--KGGKGAELAAALPDLPPVRLRLG 178
DB 73 RLSTHEDPQLIDAT-lakvpsllrslrvlkggsgalgaclhgtlttlpagisdliac 131
QY 179 LENDERAYSAPLELPIC--EATGTPLVFDAHHVHVDKLPD 217
DB 132 lahldlsfnrltptcvelhgdallishnhl--selpe 170

RESULT 11

AAAG90354
ID AAG90354 standard; Protein; 497 AA.

XX AC AAG90354;

XX DT 26-SEP-2001 (first entry)

XX DE C glutamicum protein fragment SEQ ID NO: 4108.

XX KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
organic acid synthesis.

XX OS Corynebacterium glutamicum.

XX PN EP1108790-A2.

XX PD 20-JUN-2001.

XX PF 18-DEC-2000; 2000EP-0127688.

XX PR 16-DEC-1999; 99JP-0377484.

XX PR 07-APR-2000; 2000JP-0159162.

XX PR 03-AUG-2000; 2000JP-0280988.

XX XX (KYOW) KYOWA HAKKO KOGYO KK.

XX XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A.

XX XX WPI: 2001-376931/40.

XX DR N-PSDB: AAH65573.

XX PT Novel polynucleotides derived from Corynebacterium bacteria. For identifying mutation point of a gene, measuring expression of a gene, analysing expression profile or pattern of a gene and identifying homologous gene

XX PS Claim 17; SEQ ID NO: 4108; 246pp + Sequence Listing; English.

XX CC The present invention provides a number of nucleotide and protein sequences from the Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of corynebacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived

```

PR 09-JUL-1999;          99DE-1032129.
PR 09-JUL-1999;          99DE-1032226.
PR 14-JUL-1999;          99DE-1032920.
PR 14-JUL-1999;          99DE-1032920.
PR 14-JUL-1999;          99DE-1032922.
PR 14-JUL-1999;          99DE-1032924.
PR 14-JUL-1999;          99DE-1032928.
PR 14-JUL-1999;          99DE-1032930.
PR 14-JUL-1999;          99DE-1032933.
PR 14-JUL-1999;          99DE-1032935.
PR 14-JUL-1999;          99DE-1032973.
PR 14-JUL-1999;          99DE-1033002.
PR 14-JUL-1999;          99DE-1033003.
PR 14-JUL-1999;          99DE-1033005.
PR 31-AUG-1999;          99DE-1033006.
PR 31-AUG-1999;          99DE-1041378.
PR 31-AUG-1999;          99DE-1041379.
PR 31-AUG-1999;          99DE-1041390.
PR 03-SEP-1999;          99DE-1041391.
PR                                99DE-1042088.
PA      (BADI ) BASF AG.
XX
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX WPI; 2001-061974/07.
DR N-PSDB; AAF71228.
DR
XX
PT New isolated Corynebacterium glutamicum nucleic acid for production or
PT modulation of production of fine chemicals such as amino acids,
PT nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins
PT or enzymes -
XX
PS Claim 20; Page 390-391; 712pp; English.
XX
CC AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis
CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The
CC C. glutamicum HA genes (I) can be used in vectors for expression in host
CC cells and production of fine chemicals, such as, an organic acid,
CC proteinogenic or nonproteinogenic amino acid (preferred), purine or
CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated
CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,
CC polyketide or enzyme. The amino acids produced can be lysine, glutamine,
CC glutamate, valine, aspartate, glycine, serine, threonine, methionine,
CC cysteine, alanine, leucine, isoleucine, arginine, proline, histidine,
CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can
CC be modulated. The presence of (I) or HA proteins encoded by them are
CC used for diagnosing the presence or activity of Corynebacterium
CC diphtheriae. (I) can be used to map the C. glutamicum genome or can be
CC used as markers for genetically engineered Corynebacterium or
CC brevivibacterium. The HA proteins encoded by the (I) are used to maintain
CC homeostasis in C. glutamicum or help the microorganism to adapt to
CC different environmental conditions.
XX
SQ Sequence 497 AA;
Query Match 5.9%; Score 90; DB 22; Length 497;
Best Local Similarity 23.5%; Pred. No. 0.54;
Matches 72; Conservative 31; Mismatches 90; Indels 114; Gaps 14;
QY 20 LSRYSRALSPAREAKLLDLVS-----SNIKTLRGAADYCAAHDIRLY 61
Db 68 lkswsalgaertgylkiateiessealaltntrengspisergrasnaag----if 123
QY 62 RUSSSLFFMLDAGDD-----TGAA-----VLTHAPOLLE 92
Db 124 ryfatlapwld--gedirpfagsaesivdkdpgvcaliapwnfpnlavviklapall- 180
QY 93 AGHAFTDAGVRLMHPE-----QFVLNSDRPVERESSVRAMSAHARVWDGLGARTP 145
Db 181 -----agctviikpasptpsirfiieaeagvpagvnllitgsgrfyd--alvrhp 231
QY 146 WNLLLLHGGRGGRAELAALIPDPVPRLRCLENDERAYSPAELLPCIEATGTPLVFD 205

```

Tue Jan 15 14:12:19 2002

CC used for the recombinant production of epothilones in a heterologous host
CC that is more amenable to fermentation.

XX Sequence 1421 AA;

Query Match 5.9%; Score 90; DB 21; Length 1421;
Best Local Similarity 21.5%; Pred. No. 2.6;
Matches 78; Conservative 37; Mismatches 122; Indels 126; Gaps 19;
QY 5 VCLTVGPEVFRVTLSRYRALSPAEREAKLLDLY-----SSNIKTLRGAADYCA 54
DB 129 vfqigp-----seyeaalpqtasaeidahgglgtmpsvqagrisyalgrrpcv 179
QY 55 AHDRLRLSSLPMLDLAGDDTGAATVTHLAPQLLEAGHAFT--DAGVRLLMHPEQFI 112
DB 180 avdt---aysssl-----vavhllacqslrsgcectalagvslmlspstlv 222
QY 113 VLNSDRPEVRESSVRAMSAHARVMDGLGLARTPNWLL--LHGK--GGR----- 158
DB 223 wlsktralardgrckafsaec---dgfgrgegcavvvlkrlsgradgdrilavirgsai 279
QY 159 ---GAELAALIPD-----LPDPVRLR-----L 177
DB 280 nhdgassltvongssqeqivlkradadagcaassvgvyeahgtgtltdgpieiqalnavy 339
QY 178 GLENDERAYSPAELLPICEATGTP-----LVFDAHHVHVDKLPDQE--DPSV-- 223
DB 340 gjgrd--vatpllgsvktnighpeyasgitgllkvvisiqhgqipahlhaqalnprisw 397
QY 224 ---REWVLARATWQPPEWVH-----LSNGIEGPQDRRHSHLI--ADFPsAYADVPOIEV 275
DB 398 gdlrlvtvtrtpw--pdwntpragvssfgmsg-----tnahvvleapaaatctppaper 451
QY 276 EAK 278
DB 452 pae 454

RESULT 14
AAW52844
ID AAW52844 standard; Protein; 1891 AA.

XX AC AAW52844;
XX DT 24-JUL-1998 (first entry)
XX DE Amycolatopsis mediterranei strain wt3136 5.7 kb KpnI fragment protein.
XX KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
XX KW polyketide synthase; actinomycete; ansamycin.

XX OS Amycolatopsis mediterranei.
XX PN W09807868-A1.
XX PD 26-FEB-1998.
XX PF 18-AUG-1997; 97WO-EP04495.
XX PR 20-AUG-1996; 96EP-0810551.
XX (NOVS) NOVARTIS AG.
XX PA Engel N, Schupp T, Toupet C;
XX PI WPI: 1998-169172/15.
XX DR N-PSDB; AAV21186.
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
XX to produce rifamycin and rifamycin analogues
XX Example 6; Page 43-53; 205pp; English.
XX PS

XX The present sequence represents Amycolatopsis mediterranei strain wt3136
CC 5.7 kb KpnI fragment protein, from the present invention. The present
CC invention describes a Amycolatopsis mediterranei rifamycin synthesis
CC gene cluster DNA fragment comprising a DNA region involved directly or
CC indirectly in the gene cluster responsible for rifamycin synthesis,
CC including the adjacent DNA regions to the right and left which, by
CC reason of their function in connection with rifamycin biosynthesis,
CC qualify as constituents of this rifamycin gene cluster, and functional
CC fragments, derivatives or constituents of these. The Amycolatopsis
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC for producing rifamycin, rifamycin analogues or precursors. It can also
CC be used for inactivating or modifying genes involved in ansamycin or
CC rifamycin biosynthesis. The DNA can be used for constructing mutant
CC actinomycetes strains from which the natural rifamycin or ansamycin
CC biosynthesis gene cluster has been partly or completely deleted. The
CC DNA fragment can be used for assembling a library of polyketide
CC synthases, which can be used for assembling a library of polyketides.
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.

XX Sequence 1891 AA;

Query Match 5.9%; Score 90; DB 19; Length 1891;
Best Local Similarity 22.3%; Pred. No. 4.1;
Matches 54; Conservative 21; Mismatches 61; Indels 106; Gaps 10;
QY 22 RYRALSPAERAKLLDLYSSNIKTLRGAADYCAAHDRLRLSSLPMLDLAGDDTGA 81
DB 1023 rlaglapaeqaallldvrtqv-----al 1046
QY 82 VLTHLAPQLLEAGHAFTDAGVRLLMHPEQFTVLNSDRPEVRESSVRAMSAHARVMDGLGL 141
DB 1047 vlgahgepavradtafktdgdsfslsvel-----rnlreas-----gkl 1087
QY 142 ARTPMNLLLLHGGKGGRAELALIPDLPVRLRLGLEN---DERAYSPAELLPICEAT 198
DB 1088 p-----atlvdytpvalarylrdefgdtvattptvataaada- 1126
QY 199 GTPLVFDAHHVHVDKLPDQEDPSVREWVLARATWQPPE--WQVHLSNGIEG----PQ 252
DB 1127 gepi-----aivgmactlpggvtqpegilwrlvr--dqleglsipe 1165
QY 253 DR 254
DB 1166 dr 1167

RESULT 15

AAW52849
ID AAW52849 standard; Protein; 3413 AA.

XX AC AAW52849;
XX DT 24-JUL-1998 (first entry)
XX DE A. mediterranei rifamycin synthesis gene cluster fragment protein E.
XX KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
XX KW polyketide synthase; actinomycete; ansamycin.
XX OS Amycolatopsis mediterranei.
XX PN W09807868-A1.
XX PD 26-FEB-1998.
XX PF 18-AUG-1997; 97WO-EP04495.
XX PR 20-AUG-1996; 96EP-0810551.
XX (NOVS) NOVARTIS AG.

```

XX Engel N, Schupp T, Toupet C;
PI
XX WPI; 1998-169172/15.
DR N-PSDB; AAV21187.
XX
XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT to produce rifamycin and rifamycin analogues
XX
XX Claim 6; Page 170-187; 205pp; English.
XX
XX The present sequence represents a Amycolatopsis mediterranei rifamycin
CC synthesis gene cluster ORF E protein from the present invention. The
CC DNA fragment comprises a DNA region involved directly or indirectly
CC in the gene cluster responsible for rifamycin synthesis, including
CC the adjacent DNA regions to the right and left which, by reason of
CC their function in connection with rifamycin biosynthesis, qualify
CC as constituents of this rifamycin gene cluster, and functional
CC fragments, derivatives or constituents of these. The Amycolatopsis
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC for producing rifamycin, rifamycin analogues or precursors. It can also
CC be used for inactivating or modifying genes involved in ansamycin or
CC rifamycin biosynthesis. The DNA can be used for constructing mutant
CC actinomycetes strains from which the natural rifamycin or ansamycin
CC biosynthesis gene cluster has been partly or completely deleted. The
CC DNA fragment can be used for assembling a library of polyketide
CC synthases, which can be used for assembling a library of polyketides.
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.
XX
XX Sequence 3413 AA;
SQ
Query Match 5.9%; Score 90; DB 19; Length 3413;
Best Local Similarity 22.3%; Pred. No. 9.9;
Matches 54; Conservative 21; Mismatches 61; Indels 106; Gaps 10;
QY 22 RYRALSPAEREAKLLDYSSNIKTLRGAADYCAAHDTRLVRLSSLPMLDLAGDDTGAA 81
Db 1597 riaglapaeqeallldvrtqv-----al 1620
QY 82 VLTHLAPOLLEAGHAFTDAGVRLMHPEQFIVLSNRPEVRESSVRAMSAHARVMDGLGL 141
Db 1621 vlhagpeavradtafdtdgfsitsvel-----rnlreas-----gkl 1661
QY 142 ARTPWNLILLHGKGGKRGAEALIPDLDPVRLRLGLEN---DERAYSPAELLPICAT 198
Db 1662 p-----atlvfdyptvalarylrdelgdtvattptvataaada- 1700
QY 199 GTPLVFDAAHHVVDKLPDOEDPSREWVLRARATWQPPE--WQVVHLSNGIEG----PQ 252
Db 1701 gepi-----aivgmacrlpggvtdeglwrlvr--dglelspfp 1739
QY 253 DR 254
Db 1740 dr 1741

```

Search completed: January 15, 2002, 13:50:18
Job time: 307 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:50:57 ; Search time 57.72 seconds
(without alignments)
233.532 Million cell updates/sec

Title: US-09-724-296-2_copy_230_828

Perfect score: 3112
Sequence: 1 MRLLRNTQISKRVFTL.....KRLTARKRSRKEVEDEK 599

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	140.5	4.5	733	3	US-08-725-459B-23
2	134.5	4.3	753	3	US-08-725-459B-25
3	134.5	4.3	753	3	US-08-725-459B-26
4	133.5	4.3	1588	5	PCT-US93-07261-11
5	133.5	4.3	1663	5	PCT-US93-07261-16
6	132	4.2	414	1	US-07-667-276A-4
7	130	4.2	1104	4	US-08-923-992A-4
8	128.5	4.1	769	3	US-08-725-459B-37
9	128.5	4.1	769	3	US-08-725-459B-38
10	127.5	4.1	769	3	US-08-725-459B-35
11	127.5	4.1	769	3	US-08-725-459B-36
12	127.5	4.1	769	3	US-08-725-459B-39
13	127.5	4.1	769	3	US-08-725-459B-40
14	127.5	4.1	1964	2	US-08-790-912-3
15	127.5	4.1	2052	2	US-08-790-912-2
16	127	4.1	651	1	US-08-769-309A-17
17	127	4.1	651	3	US-08-994-570-17
18	127	4.1	708	4	US-08-235-836C-76
19	127	4.1	1780	1	US-08-769-309A-5
20	127	4.1	1780	3	US-08-994-570-5
21	127	4.1	2308	1	US-08-015-973-1
22	127	4.1	2308	2	US-08-448-164-1
23	127	4.1	2308	4	US-08-081-929-2
24	125.5	4.0	680	3	US-08-725-459B-3
25	125.5	4.0	730	3	US-08-725-459B-2
26	125.5	4.0	733	3	US-08-725-459B-29
27	125.5	4.0	733	3	US-08-725-459B-30

28 125.5 4.0 773 1 US-08-524-757-6
29 125.5 4.0 773 3 US-08-725-459B-1
30 125.5 4.0 773 3 US-08-725-459B-79
31 124 4.0 1093 5 PCT-US93-03077-1
32 123.5 4.0 733 3 US-08-725-459B-22
33 123.5 4.0 770 2 US-08-209-521-13
34 123 4.0 770 4 US-08-961-810-123
35 123 4.0 770 4 US-08-352-902D-123
36 121 3.9 700 4 US-08-235-836C-66
37 121 3.9 748 3 US-08-725-459B-24
38 121 3.9 800 6 5183745-3
39 120 3.9 748 3 US-08-725-459B-27
40 119.5 3.8 694 1 US-08-339-152A-18
41 119.5 3.8 694 2 US-08-007-999B-5
42 119.5 3.8 694 2 US-08-689-276A-5
43 119.5 3.8 695 1 US-08-123-702-2
44 119.5 3.8 695 1 US-08-339-152A-30
45 119.5 3.8 695 2 US-08-104-165-1

ALIGNMENTS

RESULT 1
US-08-725-459B-23
; Sequence 23, Application US/08725459B
; Patent No. 6084068
; GENERAL INFORMATION:
; APPLICANT: CONAWAY, RONALD C.
; APPLICANT: CONAWAY, JOAN W.
; TITLE OF INVENTION: ELONGIN A AND C FUNCTIONAL DOMAINS
; NUMBER OF INVENTIONS: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SIDLEY & AUSTIN
; STREET: 717 N. HARWOOD, SUITE 3400
; CITY: DALLAS
; STATE: TX
; COUNTRY: US
; ZIP: 75201-6507
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08725,459B
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: HANSEN, EUGENIA S.
; REGISTRATION NUMBER: 31,966
; REFERENCE/DOCKET NUMBER: 11146/07501
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 214-981-3300
; TELEFAX: 214-981-3400
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 733 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..733
; OTHER INFORMATION: /note= "amino acids 1-480 and
; OTHER INFORMATION: 521-773 of Elongin A"
; US-08-725-459B-23

Query Match 4.5%; Score 140.5; DB 3; Length 733;
Best Local Similarity 20.1%; Pred. No. 0.00083;
Matches 82; Conservative 75; Mismatches 167; Indels 83; Gaps 15;

NAME: Blasdale, John H. C.
REGISTRATION NUMBER: 31,895
REFERENCE/DOCKET NUMBER: DX0288K
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-822-7398
TELEFAX: 201-822-7039
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1588 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-07261-11

Query Match 4.3%; Score 133.5; DB 5; Length 1588;
Best Local Similarity 19.4%; Pred. No. 0.011;
Matches 108; Conservative 84; Mismatches 175; Indels 191; Gaps 26;

QY 83 LKKNETELANISGPHKKSTSTSTRKARSKKATDSVSKIDESVASYDSSTHLRRSS 142
Db 7 LKQKTE-----KNEKARNALKEKKLKEQKKDAQAKADLTKE--SQDSS-----SE 51
QY 143 RS-KKPNYNSSSESEEQI-SKATKKVKOKBEEVEEVEEVEEVEEVEEVEEVEEVEE 200
Db 52 KSLKEKVNGEALKEKENKETLKKKELENOKEKEBKNKTKNNDEALANKGNKDDKKIVP 111
QY 201 BOLTPTISKRRSSAKNLEKESSTMNLDHAPREMFCDLCKPIPWGRGLGYACNLTLR 260
Db 112 KKPESVEKDLKEMELKEPEFIQ--HLKDYEEK-----EKRRNW-----ILR 152
QY 261 SMKERVFCSTRTITTTOROGLESVKQLGTONVLDLKLKVEWHNFGIHFMRVSSDLFPF 320
Db 153 SLR-----RDKLREIQ-----165
QY 321 ASHAKYGYTLFAOHLKEVGKLANKYNHRLTMHPGQVTOIASPREVVVDSAIRDAYHD 380
Db 166 ---EKLNAQLESAINELKE--RRASRRPMVMKQMGKDE-----VDEWK--KYDD 210
QY 381 EILSRMKLEQNLKDAVLIIHLGGTFEGKKE-----LDREKKNYQRLSDSVK 428
Db 211 EQAEKNGTKDEIKDK-----GDQYEEIVETKFGYMRNALGELDEVEERYE-----K 258
QY 429 AKLVLENDVSVSVDLILPLCOELNIPVLVDHWHINIYVGTILREGSLDMLPLIPIRETW 488
Db 259 KRYILK-EDGEGLDK-----VEEKELEETGYGFEKFPTRILV 296
QY 489 TRKGITQKOHVSESADPTAISG-----MKRRASDRVF-----DFPPCDPTMDL 532
Db 297 KRRKKEQKKLKEDEKELIAAEEPPDEKKIKLSDDKVVVVPVKNKSSFPD-----349
QY 533 MIEAKEQOAVFELCRRYELQN--PPCP-----LETMGPEYDQTRDGYYPGAEKRLTA- 584
Db 350 KTRAPDKKRTMF-----YRLSELFIVPRKONELAVCGDSMDSKVNG-----KKLAKST 397
QY 585 -----KRRRSKKEEVEHDE 598
Db 398 FNPFRKRRNKLKERKMOE 415

RESULT 5
PCT-US93-07261-16
Sequence 16, Application PC/TUS9307261
GENERAL INFORMATION:
TITLE OF INVENTION: P1EMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: John H. C. Blasdale
CITY: Madison
STATE: New Jersey
COUNTRY: USA
ZIP: 07940-1000

INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 753 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..753
OTHER INFORMATION: /note="amino acids 1-565 and 586-773 of Elongin A"
US-08-725-459B-26

Query Match 4.3%; Score 134.5; DB 3; Length 753;
Best Local Similarity 19.8%; Pred. No. 0.0029;
Matches 83; Conservative 75; Mismatches 173; Indels 89; Gaps 16;

QY 73 NFSFATPVDLAKKENETELANISGPHKKSTSTSTRKARSKKKATDSVSKIDESVASY 132
Db 243 SSHKEKRPVDARGDEKSSVMGREKSHKASSKEESRLL-----SEDSAKEKLPSVYVK 296
QY 133 DSSTHLRRSSRSKKPN--YNSSESESEEQISKATKKVKOKBEEVEEVEEVEEVEEVEE 190
Db 297 EXD---REGNSLKKLSPALDVASDNHFKPKKDKSEIKSKDNKQSVDSVD-----S 346
QY 191 SDEFEFPV-----VPEQLETPISKRRRS--RSSAKNLEKESTMLDDHAPR-----EMF 237
Db 347 GRGTGDLPLPRAKDKPNNLKAQEGKYRTNSDRKSPGSLPKVEEMDMODEFEQPTMSFESY 406
QY 238 DCLDKKPIPWGRGLGYACNLTLRSMKRVFCSTRTITTTIOROGLESVKOLCT-QNVLDL 296
Db 407 LSYDQP-----RKKKKVVKTSG---TALGEKGLKKDKSKSTSKNLNSA 447
QY 297 IKLVENWHNFGIHF-----RVSSDLFPF-----ASHAKYG--YTLFAQS---335
Db 448 QKLPKANENKSDKLQAPAGAEPTPRKVPDVLPAIDPLAIQTNYRPLPSLELISSFQ 507
QY 336 -----HLEVEGKLANKYNHRLTMHPGQVTOIASPREVVVDSAIRDAYHDEILSR 385
Db 508 PKRKAFSSPOEEEGAGFTGRMNSKQVYSGSKCAYLPKMMTLHQQCIRVLKNNIDSIDQ 567
QY 386 MKLNEQLN-----KDAVLIIHLGGTF-EGKKEITLDRFRKNYORLSDSVKARVLNENDOV 438
Db 568 LYRTECNHVLIEETDLQWKVCHCRDFKEERPEYESWREMYLRLQDAREQLRLTLNII 627

RESULT 4
PCT-US93-07261-11
Sequence 11, Application PC/TUS9307261
GENERAL INFORMATION:
TITLE OF INVENTION: P1EMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: John H. C. Blasdale
CITY: Madison
STATE: New Jersey
COUNTRY: USA
ZIP: 07940-1000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.5
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07261
FILING DATE: 19930805
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,531
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.5
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07261
FILING DATE: 19930805
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,531
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Blasdale, John H. C.
REGISTRATION NUMBER: 31,895
REFERENCE/DOCKET NUMBER: DX0288K
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-822-7398
TELEFAX: 201-822-7039
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1663 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
STRAIN: Malayan Camp
PCT-US93-07261-16

Query Match 4.3%; Score 133.5; DB 5; Length 1663;
Best Local Similarity 19.4%; Pred. No. 0.012;
Matches 108; Conservative 84; Mismatches 175; Indels 191; Gaps 26;
QY 83 LKNETELANISGPHKSTSTSTRKRSKKKATDSVSKIDESVASYDSSTHLRSS 142
Db 7 LKQKTE-----KNEARNALKEKQKNDKAQAKDLTKKE--SQDSS-----SE 51
QY 143 RS-KKPVNYSSESSEEQI-SKATKKVKQKEEYVEVEDEKSLKNSSSDEFEPPVP 200
Db 52 KSLKVNGEALKEKENKETLKKLENEKEKEKKNKIDNDEALKNKNDKDKKIYP 111
QY 201 EQLETPIKRRRSRSKAKLEKSTNMLEDHAPREFCDLCKPIPWGRGLGVACNTILR 260
Db 112 KKPESVEKDLKEMELKEFIQ---HLKDYEEK-----EKRRNW-----TLR 152
QY 261 SKKERVCSRTCRITFIQDGLSVKQLGTQNVLDLILKLVNHNFGIHFMRVSDLPFF 320
Db 153 SLR-----RDKLREIQ-----EKRRNW-----TLR 152
QY 321 ASHAKYGYTLFAQSHLEEVGKLANKYNHRLDMHPGOYTOIASPREVVVDSAIRLAYHD 380
Db 166 ---EKLNAQLESAINELKE--RRAARRPMVKMQRGMKDE-----VDEWIK--KYDD 210
QY 381 EILSRMKLNEQLNKDAVLIHIGGTFEGKKE-----LDPRKNYQRLSDSVK 428
Db 211 EQAENGTGDEIKDK-----GDGYEIVETKFGYMRNALGELDEYERYE-----K 258
QY 429 ARVLVENDVSVNSVODLLPLCQELNPLVDWHHNIYVPTILREGSLDMLPIPTIRETW 488
Db 259 KRYLKL-EDGEGLDK-----VEEKLTEGTGYFRKFFPTTRILV 296
QY 489 TRKGTQKHYSSESADPTAISG-----MKRRAHSRVF-----DFPPCDPTMDL 532
Db 297 KRRNKEQKRLKEDKEKLLAAEEDDEKIKLKDDSDKVVVPMKNKSSFPD----- 349
QY 533 MLEAKEQOAVFELCRRYELQN--PPCP-----LEIMGPEYDQTRDGYPPGAERLITA- 584
Db 350 KFRAPDKARTMF-----VRLSELPFIVPRKDNELAVCGDSMDSRVNG-----KKLKST 397
QY 585 -----KRRSRKEEVEDE 598
Db 398 FNPFKRRRNLKRMQE 415

RESULT 6

US-07-667-276A-4
Sequence 4, Application US/07667276A
Patent No. 5470971
GENERAL INFORMATION:
APPLICANT: Kondo, Keiji
TITLE OF INVENTION: STRESS-INDUCED PROTEINS, GENES CODING
THEREFOR, TRANSFORMED CELLS OF ORGANISMS, METHODS AND
APPLICATIONS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weiser & Associates
STREET: 230 S. Fifteenth Street, Suite 500
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/667,276A
FILING DATE: 11-MAR-1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 377,535IP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 414 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-667-276A-4

Query Match 4.2%; Score 132; DB 1; Length 414;
Best Local Similarity 26.3%; Pred. No. 0.0019;
Matches 40; Conservative 36; Mismatches 56; Indels 20; Gaps 3;
QY 51 KSLLPMSKTTLMPOVNI GANSPSAETPVDLKKENETELANISGPHKSTSTST--- 106
Db 26 KAVSSSSSSSSSS-----SSSSSESESESESESSSSSSSSSSSSSSSSSS 76
QY 107 -----RKRARRSKKATDSVSKIDESVASYDSSTHLRSSRSKPKPVNYSSESSE 161
Db 77 EATKKEESKDSSSSSSSSSDEEEEEKETPKKESSKSSSDS--SSSSSSSSSE 134
QY 162 ISKATKKVKQKEEYVEVEDEKSLKNSSSD 193
Db 135 ESNDKRRKSEDAEEDEESSNKKOKNKEETEE 166

RESULT 7

US-08-923-992A-4
Sequence 4, Application US/08923992A
Patent No. 6280738
GENERAL INFORMATION:
APPLICANT: Tai, Joseph Y.
TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,992A
FILING DATE: 05-SEP-1997
CLASSIFICATION: 536
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/024,707
FILING DATE: 06-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1104 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-923-992A-4

Query Match 4.2%; Score 130; DB 4; Length 1104;
Best Local Similarity 19.7%; Pred. No. 0.013;
Matches 119; Conservative 88; Mismatches 187; Indels 210; Gaps 29;
QY 44 HCLPTLSKLLPMSKTTLSMLPQVNICANSFSAETPVDLKKENETELANISGPHKKS-- 101
DB 142 HNKPNQOKTKOSDFKVDLS-----NIDKELNHQKSOV---EAMAEQAGITNEDKRDMSL 192
QY 102 -TSTTRKARSSKKKATDSYSDKIDESVASYDSSTHLRRSSRSKPKPVNNSSESSES-- 158
DB 193 KIEDIRKQQAQADKKEAEV--KYRELKGLFSSTKAGLDQOIOEHVKKETSSEENTQK 250
QY 159 -EQISKATKKVKQKEE-----EYVEEVDKSLKNSSSDEFEPPVPEQLETPISK 209
DB 251 VDEHYANSLOLQAQSLSELDKATTNEQATQVKNQFLENQAQKLEIOPLIK---ETNVKL 307
QY 210 RRRSRSAKNLEKE---STMNLDHAPREMFDCLDKPI--PWGRGLGYACLNITLRSKM 263
DB 308 YKAMESLEQVEKELKHNSANLQDLVAK-----SKELVREYEGKLN----- 349
QY 264 ERVFCSTCRITTIQDGLGVESVQKLTQNVLDLILKEVNHNFHFMVSSDLPFPASH 323
DB 350 -----QSKNLPKQLQEEEAHSLKQVVE-----HFRK----- 377
QY 324 AKYGYTLEFAQSHLEEVGKLANKYNIHRLTHMPCQYQTASPREVVDSALBOLAYHDEIL 383
DB 378 -----KFKTS---EQV-----TPKKRVK-----RDLAANEN-- 400
QY 384 SRMKLNQKNAQVLIHLGGTFEGK-----KETL-----DRFR 417
DB 401 NQKIETLVSPENITV-----YEGEDVKTFTAKSDSKTTLDFSLTLTKYNPSVSDRIS 454
QY 418 NQYRLSDSYK--ARLVLENDVDSWSVODLLPLCOELNIPLDVLDHNNHIVPG-----TLRE 472
DB 455 TNYKTNTDNHIAEITKN-----LKNQSQTVTLKARDSDGNVVEKTFITVQK 504
QY 473 GSLDMLPIPTIETWTRKGIQKQHYSESADPTAISGKMRRAHSDKRVDFPFCPTMDL 532
DB 505 KEERQVPTPTPEQKHSKTEQNVPOE---PKSNQKNQLOELIKSAQO----- 546

QY 533 MIEAKKEQAOVFLCRRRYELQNPQPCPLEIMGPYDQTRGDYPPGAERKRLTARKRRSRKE 592
DB 547 --ELEKLEKAIKEL-----MEQPEIP---SNPEY-----GLOKSLW-----ESOKE 582
QY 593 EVEE 596
DB 583 PIQE 586
RESULT 8
US-08-725-459B-37
Sequence 37, Application US/08725459B
Patent No. 6084068
GENERAL INFORMATION:
APPLICANT: CONAWAY, RONALD C.
APPLICANT: CONAWAY, JOAN W.
TITLE OF INVENTION: ELONGIN A AND C FUNCTIONAL DOMAINS
NUMBER OF SEQUENCES: 79
CORRESPONDENCE ADDRESS:
ADDRESSEE: STIDLEY & AUSTIN
STREET: 717 N. HARWOOD, SUITE 3400
CITY: DALLAS
STATE: TX
COUNTRY: US
ZIP: 75201-6507
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,459B
FILING DATE: 04-OCT-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: HANSEN, EUGENIA S.
REGISTRATION NUMBER: 31,966
REFERENCE/DOCKET NUMBER: 11146/07501
TELECOMMUNICATION INFORMATION:
TELEPHONE: 214-981-3300
TELEFAX: 214-981-3400
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: protein
LOCATION: 1..769
OTHER INFORMATION: /note= "amino acids 1-552 and
OTHER INFORMATION: 557-773 of Elongin A"
US-08-725-459B-37

Query Match 4.1%; Score 128.5; DB 3; Length 769;
Best Local Similarity 20.0%; Pred. No. 0.01;
Matches 88; Conservative 72; Mismatches 168; Indels 111; Gaps 18;
QY 73 NPSAETPVDLKKENETELANISGPHKKSSTSTRKRARSSKKKATDSVSDKIDESVASY 132
DB 243 SSHKRRPVDVADGDEKSSVNGREKSHKASSKEESRLL-----SEDSAKEKLPSSVVK 296
QY 133 DSSHTLRRSSRSKPKPVN--YNSSESSEEQISKATKKVKQKEEEYVEEVDKESLKNES 190
DB 297 EKD--REGNSLKKKLSPALDVASDNHFKPKHKDSEKIKSDKNQKQSDVD-----S 346
QY 191 SSDEFEPV-----VPEQLFTPISKRRS--RSSAKNLEKSTMNLDHAPR-----EMF 237
DB 347 GRUTGDLPLRAKDKVFNLLKAQEGKVRTNSDRKSPGSLPKVEEMDMDDFEQPTMSFESY 406

```

QY 238 DCLDKPIPWGRGLGYACLTILRSMKERVFCSTRTITTIQDGLSVKOLGT-QNVLDL 296
Db 407 LSYDQP-----RKKKKVKTSG---TALGEKGLKKKDKSKTKNLNSA 447
QY 297 IKLVENHNFGIHF-----RVSSDLFPF-----ASHAKYG--YTFLEFAQS-- 335
Db 448 QKLPRANENKSDKLQAGAEPTPRKVPDVLPAIPAIQTNVRLPLSLELISSFO 507
QY 336 -----HLEEVGKLANKYNHR-----LTMHPQYTOIASPREV--- 367
Db 508 PKKAFSSPQEEBAGFTGRMRMSKMOVYSGSKCAYLPKMMTILHQLKNNIDSIPEVGGV 567
QY 368 -----VVDSAIRDLAYHDEILSRMKLNQKDAVLIHLGGTF-EGKKTETLDRFRKN 419
Db 568 PYSVLEPVLRCPTDQLYRIEENHVLIEE---TDQLWKVCHRDKFERPEEYESWREM 624
QY 420 YORLSDSVKARLVLENDV 438
Db 625 YLRLODAREQRLRLTNNI 643

RESULT 9
US-08-725-459B-38
; Sequence 38, Application US/08725459B
; Patent No. 6084068
; GENERAL INFORMATION:
; APPLICANT: CONAWAY, RONALD C.
; APPLICANT: CONAWAY, JOAN W.
; TITLE OF INVENTION: ELONGIN A AND C FUNCTIONAL DOMAINS
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SIDLEY & AUSTIN
; STREET: 717 N. HARWOOD, SUITE 3400
; CITY: DALLAS
; STATE: TX
; COUNTRY: US
; ZIP: 75201-6507
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/725,459B
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: HANSEN, EUGENIA S.
; REGISTRATION NUMBER: 31,966
; REFERENCE/DOCKET NUMBER: 11146/07501
; TELEPHONE: 214-981-3300
; TELEFAX: 214-981-3400
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: protein
; LOCATION: 1..769
; OTHER INFORMATION: /note="amino acid 1-556 and
; OTHER INFORMATION: 561-773 of Elongin A"
US-08-725-459B-38

```

```

Query Match 4.1%; Score 128.5; DB 3; Length 769;
Best Local Similarity 20.0%; Pred. No. 0.01;
Matches 88; Conservative 72; Mismatches 168; Indels 111; Gaps 18;
QY 73 NSFAETPVDLKKENETELANISGPHKKTSTSTRKRARRSKKATDSVSKIDESVASY 132

```

```

Db 243 SSHKKEKRPVDARGDEKSSVMGRKSKSHKASSKEESRLL-----SEDSAKELPSSVYKK 296
QY 133 DSSTHLRRSSRSKKPVN--YNSSESSESEEQISKATKKVKQKEEEYVEEVEDEKSLKNES 190
Db 297 EKD---REGNSLRKKLSPALDVAENHFKPKHKHDKSEIKSKDNKQSVDSVD-----S 346
QY 191 SDEFEFV-----VPEQLETPISKRRRS--RSSAKNLEKSTMNLDHAPR-----EMF 237
Db 347 GRGTGDPPLPRAKDKVFNLLKAQEGKVRTNSDRKSPGSLPKVEEMDDDEFPOTMFSFESY 406
QY 238 DCLDKPIPWGRGLGYACLTILRSMKERVFCSTRTITTIQDGLSVKOLGT-QNVLDL 296
Db 407 LSYDQP-----RKKKKVKTSG---TALGEKGLKKKDKSKTKNLNSA 447
QY 297 IKLVENHNFGIHF-----RVSSDLFPF-----ASHAKYG--YTFLEFAQS-- 335
Db 448 QKLPRANENKSDKLQAGAEPTPRKVPDVLPAIPAIQTNVRLPLSLELISSFO 507
QY 336 -----HLEEVGKLANKYNHR-----LTMHPQYTOIASPREV--- 367
Db 508 PKKAFSSPQEEBAGFTGRMRMSKMOVYSGSKCAYLPKMMTILHQLKNNIDSIPEVGGV 567
QY 368 -----VVDSAIRDLAYHDEILSRMKLNQKDAVLIHLGGTF-EGKKTETLDRFRKN 419
Db 568 PYSVLEPVLRCPTDQLYRIEENHVLIEE---TDQLWKVCHRDKFERPEEYESWREM 624
QY 420 YORLSDSVKARLVLENDV 438
Db 625 YLRLODAREQRLRLTNNI 643

RESULT 10
US-08-725-459B-35
; Sequence 35, Application US/08725459B
; Patent No. 6084068
; GENERAL INFORMATION:
; APPLICANT: CONAWAY, RONALD C.
; APPLICANT: CONAWAY, JOAN W.
; TITLE OF INVENTION: ELONGIN A AND C FUNCTIONAL DOMAINS
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SIDLEY & AUSTIN
; STREET: 717 N. HARWOOD, SUITE 3400
; CITY: DALLAS
; STATE: TX
; COUNTRY: US
; ZIP: 75201-6507
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/725,459B
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: HANSEN, EUGENIA S.
; REGISTRATION NUMBER: 31,966
; REFERENCE/DOCKET NUMBER: 11146/07501
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 214-981-3300
; TELEFAX: 214-981-3400
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein

```

```

LOCATION: 1..769
OTHER INFORMATION: /note= "amino acids 1-544 and
OTHER INFORMATION: 549-773 of Elongin A"
US-08-725-459B-35

Query Match      4.1%; Score 127.5; DB 3; Length 769;
Best Local Similarity 19.6%; Pred.No. 0.012;
Matches 86; Conservative 74; Mismatches 168; Indels 111; Gaps 18;

QY 73 NSFSAETPVDLKKNETELANISGPHKKTSTSTRKRARRSKKKATDSVSDKDIESVASY 132
    : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 243 SSHKEKRPVDARGDEKSSVMGREKSHKASSKEESRLL-----SEDSAKELPSSVVKK 296

QY 133 DSSTHLRRSSRSKPVN--YNSSSESSEEQISKATKYVKQEKEEYVEEVDEKSILKNES 190
    : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 297 EXD---REGNSLKKLPALDVASDNHFKPKHKDKSEKIKSDKNKGSVDSD-----S 346

QY 191 SDEFEPV-----VPQLETPISKRRRS--RSSAKNLEKSTEMNLDDHAPR-----EMF 237
    : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 347 GRGTCDPLPRADKVPNNLKQAEGKVRTNSDRKSPGLPAKVEEDMDDEFEOPTMSFSFY 406

QY 238 DCLDKPIPWGRGLGYACLTILRSMKERVCFSRTCRTITTIQRGLGESVKOLGT-QNVLDL 296
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 407 LSYDQP-----RKKKKKVVTSG---TALGEGLKKKKDKSKTSKNLSA 447

QY 297 IKLVENHNFGTHFM-----RVSSDLFPF-----ASHAKYG--YTLEFAQS-- 335
    || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 448 QKLPRKANENSKDLPAGAEPTPRKVPDVLPALDPILPAIOTNYRPLPSLELISSFQ 507

QY 336 -----HLEFEVKLANKYNHRLTMHPGO---YTQI----- 361
    ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 508 PKRAFFSPQEEEEEAGFTGRRNMNSKMQYSKSCAYLTLHQOCIRVLKNNIDSI FEVGV 567

QY 362 -ASPREVWDSAIRDLAYHEILSRMLNEQLNKDAVLIIHLGGTF-EGKETLDRFRKN 419
    ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 568 PYSVLEPVLERTCPDQLYRIECNVHLEE---TDQLMKVCHRDFFKEERPEEYESWREM 624

QY 420 YQLSDSVKARLVENDDV 438
    ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB 625 YLRQDAREQRRLTLNNI 643

RESULT 11
US-08-725-459B-36
; Sequence 36, Application US/08725459B
; Patent No. 6084068
; GENERAL INFORMATION:
; APPLICANT: CONAWAY, RONALD C.
; APPLICANT: CONAWAY, JOAN W.
; TITLE OF INVENTION: ELONGIN A AND C FUNCTIONAL DOMAINS
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SIDLEY & AUSTIN
; STREET: 717 N. HARWOOD, SUITE 3400
; CITY: DALLAS
; STATE: TX
; COUNTRY: US
; ZIP: 75201-6507
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/725,459B
; FILING DATE: 04-OCT-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: HANSEN, EUGENIA S.
; REGISTRATION NUMBER: 31,966
; REFERENCE/DOCKET NUMBER: 11146/07501
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 214-981-3300
; TELEFAX: 214-981-3400
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..769
; OTHER INFORMATION: /note= "amino acids 1-548 and
; OTHER INFORMATION: 553-773 of Elongin A"
; US-08-725-459B-36

Query Match 4.1%; Score 127.5; DB 3; Length 769;
Best Local Similarity 19.6%; Pred. No. 0.012;
Matches 86; Conservative 75; Mismatches 167; Indels 111; Gaps

QY 73 NGSFAPTVDLKKENETELANISGPHKSTSTSTPKRARSSKKKATDSVSKIDESVASV 132
   :| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 243 SSHKEKPVPDARGDEKSSVMGRKSHKASSKRLL-----SEDSAKELPSSVVKK 296

QY 133 DSSTHLRRSRSGKKPVN--YNSSSESSESEQISKATKKVKQKEEVEEVDEKSLKNES 190
   : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 297 EKD--REGNSLKKLSPALDVASDNHFHKPKHDKSEKIKSUKNKQSVDSD-----S 346

QY 191 SSDEFEPV-----VPEQLETPLSKRRS--RSSAKNLEKESTMNLDDHAPR-----EMF 237
   :| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 347 GRCTGDPLPRAKDQKVNNLKAQEGVRTNSDRKSPGSLPKVEEMDMDFEEOPTMSFESV 406

QY 238 DCLDKLPWRGLRGYACLTILTKSMKRVFCRSRTCITTIQRDGLLESVKQLGT-QNVLDL 296
   :| :| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 407 LSTDQP-----RKKKKKVVKTSG---TALGEKGLKKDKSKTSKLNLSA 447

QY 297 IKLVEWNHNFGIHF-----RVSSDLPPF-----ASHAKYG--YTLEFAQS-- 335
   || : | | | | | | | | | | | | | | | | | | | | | | | | |
Db 448 QKLPAKANENKSDKLOPAGAPTRPRKVPITDVLPALDPPLPATQTNYPRLPSLELISSFQ 507

QY 336 -----HLEVCKLANKYNHRLTMHPG-----QVTQTA----- 362
   || | | | | | | | | | | | | | | | | | | | | | | | | |
Db 508 PKRAFSSPOEEEEAGFTGRMKSMMQVYSGSKCAVLPKMQCIRVLKKNIDSIFEVGCV 567

QY 363 --SPREVVDSATDLAYHBEILLRMKLNOLNKDAVLIITHLGTF-ECKKETLDRFRKN 419
   : | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 568 PYSVLPVLERCTPDOLYRIECNHVILIE---TDQLWKVCHRDPFKEERPEYESWREM 624

QY 420 YORLSDSVKARLVLENDV 438
   || | | | | | | | | | | | | | | | | | | | | | | | | |
Db 625 YLRQDAREQRLRLTTNI 643
   || | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
US-08-725-459B-39
; Sequence 39, Application US/08725459B
; Patent No. 6084068
; GENERAL INFORMATION:
; APPLICANT: CONAWAY, RONALD C.
; APPLICANT: CONAWAY, JOAN W.
; TITLE OF INVENTION: ELONGIN A AND C FUNCTIONAL DOMAINS
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESS: SIDLEY & AUSTIN
; STREET: 717 N. HARWOOD, SUITE 3400
; CITY: DALLAS
; STATE: TX
; COUNTRY: US
; ZIP: 75201-6507
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

RESULT 12

```

US-08-725-459B-39
:
:   Sequence 39, Application US/08725459B
:
:   Patent No. 6084068
:
:   GENERAL INFORMATION:
:
:   APPLICANT: CONAWAY, RONALD C.
:
:   APPLICANT: CONAWAY, JOAN W.
:
:   TITLE OF INVENTION: ELONGIN A AND C
:
:   NUMBER OF SEQUENCES: 79
:
:   CORRESPONDENCE ADDRESS:
:
:   ADDRESSEE: SIDLEY & AUSTIN
:
:   STREET: 717 N. HARWOOD, SUITE 3400
:
:   CITY: DALLAS
:
:   STATE: TX
:
:   COUNTRY: US
:
:   ZIP: 75201-6507
:
:   COMPUTER READABLE FORM:
:
:   MEDIUM TYPE: Floppy disk
:
:   COMPUTER: IBM PC compatible
:

```


Tue Jan 15 14:12:08 2002

Db 625 YLRQDAREQRLRLTNNI 643

RESULT 14

US-08-790-912-3
; Sequence 3, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,939
; FILING DATE: 23-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 7600-401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1964 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-790-912-3

Query Match 4.1%; Score 127.5; DB 2; Length 1964;

Best Local Similarity 18.9%; Pred. No. 0.053;

Matches 86; Conservative 79; Mismatches 180; Indels 109; Gaps 18;

Qy 47 PDTLKSLPMSKTTLSMLPQVNIAGNSFSAETPVDLAKKE---NETELANIS-----G 96
Db 459 PDTVVS--DKGEFOVAPLPEYKGNTEOVKPTPEVKTKEQGPKEETEEVPVKPTETPVN 516
Qy 97 PHKKTSTSTRKRRSRKSKKATDSV--SKIDESVASYDSSTHLRBSRSRKKPVNYSNS 154
Db 517 PNEGTTEGTSTQAEAPVQPAEESTTNSEKVSPTDSSENTGE-----VSSN 562
Qy 155 ESEEEQISKATK----KVKQKEEEYVEEV---DEKSLKNFSSSDPEFPVVP-EQLET 205
Db 563 PSDSTTSVGSNKPENHNSKNSKTEVEVPVNPNEGTVGTSNQETKPVQPAEETQT 622
Qy 206 PISK-RRRSRSKAKNLEKSTMTLDHAPFDCLDKPIPWGRGLYACLTILRSKME 264
Db 623 NSGKIANENTGEVSNKPSDSKPPVEESNQEKNGATKP-----HNFCHFM 311
Qy 265 RVFCSTCRITTTQRODGLSVKOLGTONVLDLKLVEWN-----HNFCHFM 311
Db 662 -----ENSGNTTSNGQTEPEKKLELRNVDIELYSQTNGTYRQHVSLDGPENTDTYEV 716

Qy 312 RVSSDLF-----PFAS-----HAKYGYTLFEAQSHLEVGKLANKYNHRLTMHPGOVQT 360
Db 717 KVKSSAFKDWIPIVASITTEKRNQGSVYKITAQAKLQQ--ELENKYVDNFSFY----- 768
Qy 361 IASPREVVVDSAJRDLAYIDEILSRMKLNEQNKDAVLIHLGGTTEGKKTLDREKRY 420
Db 769 -----LDKKAKE--ENTNFTSFSNLVKAQNPNPSTGYTHLAASLANEVELGPDERSY 818
Qy 421 QRLSDSVKARLVLENDVDSVODL-LPICQELN 453
Db 819 --IKDTFTGLIGKDGKIYAIYNLKKPLFENLS 850

RESULT 15

US-08-790-912-2
; Sequence 2, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,939
; FILING DATE: 23-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 7600-401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2052 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-790-912-2

Query Match 4.1%; Score 127.5; DB 2; Length 2052;

Best Local Similarity 18.9%; Pred. No. 0.057;

Matches 86; Conservative 79; Mismatches 180; Indels 109; Gaps 18;

Qy 47 PDTLKSLPMSKTTLSMLPQVNIAGNSFSAETPVDLAKKE---NETELANIS-----G 96
Db 532 PDTVVS--DKGEFOVAPLPEYKGNTEOVKPTPEVKTKEQGPKEETEEVPVKPTETPVN 589
Qy 97 PHKKTSTSTRKRRSRKSKKATDSV--SKIDESVASYDSSTHLRBSRSRKKPVNYSNS 154
Db 590 PNEGTTEGTSTQAEAPVQPAEESTTNSEKVSPTDSSENTGE-----VSSN 635
Qy 155 ESEEEQISKATK----KVKQKEEEYVEEV---DEKSLKNFSSSDPEFPVVP-EQLET 205

```

Db 636 PSDSTTSVGSNKPEHNDKNSKENSEKTVEEVVNPNEGTVEGTSNQETKFPVOPAEETQT 695
QY 206 PISK-RRRSRSKALEKSTWNLDDHAPRPMFCCLDKPIWRCRLGYACLNILRSMKE 264
Db 696 NSGKTANENTGEVSNKPSDSKPPVEESNOPEKNGTATKP----- 734
QY 265 RVFCSTCRITTIORDGLESVKQLGTQNVLDLIKLVEMN-----HNFQIHEM 311
Db 735 -----ENSGNTTSENGQTEPEKKLELRNVSDIELYSQTNCTYRQHVSLDGIPTDTYFV 789
QY 312 RVSSDLF-----PTAS-----HAKYGYTLEFAQSHLEEVGKLANKYNHRLTMHPCQYQ 360
Db 790 KVKSFAKDVYIPVASITEEKRNQSVYKITAKAEKLOQ--ELENKYVDNFSFY----- 841
QY 361 IASPREVVVDSAIRDLAYHDEILSRMKLNQNLKNDKAVLIIHLGGTFEGKKEITLDRFRKNY 420
Db 842 -----LDKAKE--ENTNFTSFSNLVKAINONPSGTYHLAASLNANEVELGPDERSY 891
QY 421 ORLSDSVKARLVLENDVSVSWODL-LPLCOELN 453
Db 892 --IKDTETGRLIGKDGKIYAIYNLKKPLFENLS 923

```

Search completed: January 15, 2002, 13:51:01
 Job time: 330 sec

Db 257 RSHADYVDANFLPLERFRQWGTNIDFMIEAKQKQKALLRL 298

RESULT 2
GOG4_HUMAN STANDARD; PRT; 2230 AA.
ID GOG4_HUMAN
AC Q13439; Q14436; Q13270; Q13654;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GOLGI AUTOANTIGEN, GOLGIN SUBFAMILY A 4 (TRANS-GOLGI P230) (256 KDA
DE GOLGIN) (GOLGIN-245) (72.1 PROTEIN).
GN GOLGA4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96215236; PubMed=8626529;
RA Erlich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.,
RT "Molecular characterization of trans-Golgi p230: a human peripheral
RT membrane protein encoded by a gene on chromosome 6p12-22 contains
RT extensive coiled-coil alpha-helical domains and a granin motif.";
RL J. Biol. Chem. 271:8328-8337(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Seelig H.P.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 131-2230 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=96125112; PubMed=8537393;
RA Fritzier M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;
RT "Molecular characterization of golgin-245, a novel Golgi complex
RT protein containing a granin signature.";
RL J. Biol. Chem. 270:31262-31268(1995).
RN [4]
RP SEQUENCE OF 524-672 FROM N.A.
RC TISSUE=Gastric fundus;
RA Balague C.;
RL Thesis (1994), Instituto municipal de investigacion medica, Spain.
-!- FUNCTION: MAY PLAY A ROLE IN VESICULAR TRANSPORT FROM THE TRANS-
GOLGI.
-!- SUBCELLULAR LOCATION: ASSOCIATED WITH THE CYTOPLASMIC FACE OF THE
GOLGI MEMBRANE.
-!- ALTERNATIVE PRODUCTS: DIFFERENTS ISOFORMS MAY ARISE BY ALTERNATIVE
SPLICING.
-!- DISEASE: ANTIGEN IN THE AUTOIMMUNE DISEASE SJOEGREN'S SYNDROME AND
IN HEPATITIS B.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U41740; AAC50434.1; -
EMBL; X82834; CAA58041.1; -
EMBL; U31906; AAC51791.1; -
EMBL; X76942; CAA54261.1; -
MIM; 602509; -
InterPro: IPR00237; GRIP.
DR Pfam; PF01465; GRIP; 1.
KW Golgi stack; Antigen; Coiled
FT DOMAIN 133 237 COILED COIL (POTENTIAL).
FT DOMAIN 276 1011 COILED COIL (POTENTIAL).
FT DOMAIN 1033 1214 COILED COIL (POTENTIAL).
FT DOMAIN 1259 2152 COILED COIL (POTENTIAL).
FT VARSPLIC 2154 2185 TPYKGNLYHTDVSLEGEETFEYLRKVLFEY -> HLTKV
AICTRMSHSLNPLNLSICEKCFUSI (IN ISOFORM

FT VARSPLIC 2186 2230 MISSING (IN ISOFORM 2).
FT VARSPLIC 2103 2109 MISSING (IN ISOFORM 3).
FT VARSPLIC 2222 2230 FTSPRSGLF -> SWLRSS (IN ISOFORM 4).
FT CONFLICT 188 188 R -> K (IN REF. 3).
FT CONFLICT 220 220 Y -> H (IN REF. 3).
FT CONFLICT 276 276 T -> A (IN REF. 3).
FT CONFLICT 584 584 K -> E (IN REF. 3).
FT CONFLICT 628 628 T -> A (IN REF. 3).
FT CONFLICT 630 630 K -> E (IN REF. 3).
FT CONFLICT 682 682 K -> N (IN REF. 3).
SQ SEQUENCE 2230 AA; 261139 MW; 3BB733DB1EA86134 CRC64;

Query Match 5.1%; Score 160; DB 1; Length 2230;
Best Local Similarity 19.4%; Pred. No. 0.46;
Matches 122; Conservative 109; Mismatches 233; Indels 164; Gaps 27;

QY 41 SRFHCLPDTLKSL---LPMSSKTTLSMLPQVNIAGNSFSAETP-VD-LKKENETELANIS 95
Db 1451 SRFTHQNTVKELOIQLELKSKEAYEKDQINLLAEELDQONKRFDCLEGGEDDKSKM- 1509
QY 96 GPHKSTSTSTKRAKSSKKKATDSVSKIDESVASYSTHLRRSSSKKPKVNVNSSSE 155
Db 1510 --EKESNLETELKQSTAR--IMELEDHITOKTIEISLNEVLK-----NTNQOKD 1556
QY 156 SESEOTSATKKVKQKEEE--YVEEVDEK--SLKNSSSDSEFEFVVPVPELFTPTSKRRR 212
Db 1557 IEHKELVOKLOHFOELGCEKDNVKEAEKILTLENQVYS-----MKALETKKKELEH 1610
QY 213 SRSSAKNEKESTMNLDDHAPREMFCCLDKPIWRGRGLYACLNLTILRSMKERVCSRTC 272
Db 1611 VNLVSKSKEE-----LKALEDRLSESAA 1635
QY 273 RITTIORDGLESV---KOL-----GTQNVLDLILKLVENHNHFIHFMR--- 312
Db 1636 KLAELURKAEOKIAIKKOLLSSOMEKEKEQYKKGTSHELSELNTKLOEREREVHILEEKL 1695
QY 313 --VSSD-----LFPFASHAKYGYTLLEFAQSHLEEVGKLANKYNHRLTMHPGOYTO---- 360
Db 1696 KSVSSQSETLIVPRSAKNVAAYT---EQEADSGCVQKTYEEKISVLORNLKEKLL 1752
QY 361 --IASPREVVDSATRDLAYHDEILSRMKLNE-QLNKDAVLIILHGGTFEGKKTLDLRF 417
Db 1753 QRVGQKEETVSHFEMRCQYQERLIKLEHAEAKOHEQDSMIGHL-----QEELEKN 1805
QY 418 KNVQRL-----SDSVKARLVLEN--DDVSVSVDLLPLCO-----ELNIPVL 458
Db 1806 KKYSLIVAGHVEKEGKNNIQAKONLENVDDVQKTLQEKELTCQILEQKIKELDSCLVR 1865
QY 459 DWHHINIVPGLTIR---EGSLDMLPIPTIRETWTRKGTOKOHYSEADPTAISGMKRR 515
Db 1866 QKEVIRVEMEELTSYKEQLAQQQMDGRNKPTELEENTEESKSHLVQPKLLSMEAQ- 1924
QY 516 HSDRVDFDPCCPTDMLMEAKEQAVFELCR-----RYELQNPCCPLEIMGPYD 567
Db 1925 HNDLEFKAGAE-----REKOKLKEIVRLQKDLMLRKEHQO---ELEILKKEYD 1972
QY 568 QTRDQYPPGAEKRLTARKRSRKEEVE 595
Db 1973 QERE-----EKIQEQEEDLE 1987

RESULT 3
GAR2_SCHPO
ID GAR2_SCHPO STANDARD; PRT; 500 AA.
AC P41891; O13707;
DT 01-NOV-1995 (Rel. 32, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN GAR2.
GN GAR2 OR SPAC140.02.
OS Schizosaccharomyces pombe (Fission yeast).


```
QY 494 TQKHYESADPTALSGMKRRRAHSDRVFD 522
DB 619 AQDIYWDETAKGTPTIS-----IERDITVFD 642

RESULT 5
RRPP_PILHE STANDARD; PRT; 568 AA.
AC P32532;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RNA POLYMERASE ALPHA SUBUNIT (EC 2.7.7.48) (NUCLEOCAPSID
DE PHOSPHOPROTEIN).
GN P.
OS Human parainfluenza 1 virus (strain CI-14/83).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Paramyxovirus.
OX NCBI_TaxID=31606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295573; PubMed=1318610;
RA Power U.F., Ryan K.W., Portner A.;
RT "The P genes of human parainfluenza virus type 1 clinical isolates
RT are polyclonal and microheterogeneous.";
RL Virology 189:340-343(1992).
CC -!- FUNCTION: THIS PROTEIN IS PROBABLY A COMPONENT OF THE ACTIVE
CC POLYMERASE. IT MAY FUNCTION IN TEMPLATE BINDING.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M74080; AAA46830.1; -
DR PIR; E40234; RNNZ83.
DR InterPro; IPR002693; Paramyxo-P.
DR Pfam; PF01806; Paramyxo-P; 1.
KW Transferase; RNA-directed RNA polymerase; Nucleocapsid;
KW Phosphorylation.
SQ SEQUENCE 568 AA; 64586 MW; 8795D10C1CD3CCD7 CRC64;

Query Match 4.7%; Score 147; DB 1; Length 568;
Best Local Similarity 21.2%; Pred. No. 0.41; Indels 122; Gaps 23;
Matches 114; Conservative 87; Mismatches 214;

QY 62 LMLPQVNIANSFSAETPVDLKKENETELANISGPHKKTSTSTRKARRSKKATDSV 121
DB 36 LSYKP-TEIGEDRWLHNI IDNPKENKSKSDDDNNKRAISTQDH-RSSEE---SGI 90
QY 122 SDKIDESVSDSTHLR-----RSSR---SKKPVNNSSE-----SESEE 160
DB 91 SRRTGES-----KTETHARLDQQGIHRSRGSTPNLPENMGDERNTRIDESPNRRH 146
QY 161 QISKATKKVKO-----KEEPEYVEEDEKSLKNSSSDEFEPVVPQLETPISKRRSR 214
DB 147 QRSVPTDEDRKWAENSKREEDQVEGPEEVGSGTSLSDGGE-----GRTN 192
QY 215 SSANKLEKSTMNLDIAPRMFDCLDKPIPRWGRIGYACINTILSKMRKRVCSRTCRI 274
DB 193 NNGSRMETSST-----HSTR-ITDVTNPS-----ELEPAVLQKRRRP 231
QY 275 TTIOEDGLGSKVLGTOVNLDLIKLVW-NHNFQIHFMRYSSDLFFPASHAKYGYTLEFA 333
DB 232 TTKRNRTRSR---TON-SELHKSTSGDSNLENDHNTKTSQKIPP----- 273
QY 334 OSHLEEYVKLANKYNNRLTWHPGQYTOIASPREVVVDSATRDILAYHDEILSRMKLINEQLN 393
DB 274 SKNEEPAVTKNNHNRKTKHT---TNNANNARCLPTPEHDTTSNDEGTSNTSDVEM-- 328
```

```
QY 394 KOAVLIHLG-----GTFEKKETLDORFK-----NYORLSDSYKARLVLENDVSVS 442
DB 329 --AKLLVSLGVNKSQHEFELSSRASHQFAKRLKSNYKEMTFNLGMLLSVEKSLGNKV 386
QY 443 QDLLPLCQELNIPVLVDWHHHHNVPGTLESGSLDMLPLIPTIRETWTTRKGITOKOHYES 502
DB 387 EENRTLLKIQEIDSSRLHKRFSEYQKEQNSLMANLSTLHIITDRGKT-----GDP 441
QY 503 ADPTATSGMKRRRAHSDRV-----FDEPPCDPTMDLIEAKEKEQAVFELCRRRYELQNP 555
DB 442 SDTTRSPSVFTGKDNKVKYKTRF-----DPSMEALGGQEFKPDLLREDELREDIRNP 493

RESULT 6
IF38_SCHPO STANDARD; PRT; 918 AA.
AC 014164; P78791; Q9P6P4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 93 KDA SUBUNIT
DE (EIF3 P93).
GN TIF33 OR SPAC4A8.16C OR SPAC823.01C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetaceae; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PR745;
RX MEDLINE=98162722; PubMed=9501991;
RA Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;
RT "Identification of open reading frames in Schizosaccharomycetes pombe
RT cDNAs.";
RL DNA Res. 4:363-369(1997).
CC -!- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
CC METHIONYL-TRNAI AND MRNA (BY SIMILARITY).
CC -!- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE EIF3S8 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL355013; CAB90146.1; -
DR EMBL; 298762; CAB11485.1; -
DR EMBL; D89140; BAAL3802.1; -
DR InterPro; IPR000717; PCI.
DR Pfam; PF01399; PCI; 1.
DR SMART; SM00088; PINT; 2.
KW Initiation factor; Protein biosynthesis.
KW DOMAIN 1 188 ASP/GLU/SER-RICH.
FT CONFLICT 683 683 M -> I (IN REF. 2).
SQ SEQUENCE 918 AA; 104367 MW; 43B70599409688EF CRC64;
```

Query Match 4.7%; Score 145.5; DB 1; Length 918;
Best Local Similarity 23.5%; Pred. No. 0.9;
Matches 103; Conservative 57; Mismatches 172; Indels 107; Gaps 20;

```
QY 71 GANSFSAETPVDLKKNETELANISGPHKKSTSTSTRKARRSKKKKATDSVSKIDESA 130
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7 GGSDDSAESVDSSEEN-----RLTSSRLKKQDDSSSESESESESESESESESESE 61

QY 131 SYDSSTHLRRSRSSKKPVYNSSSESESESESESESESESESESESESESESESESE 190
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 SESEVEVPK-----KVAASESESESESESESESESESESESESESESESESESESE 112

QY 191 SSDFEPVVPOLPTPIKRRSSSSAKNLEKE-----STNLDHAPREM 236
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 113 ESESESESESESESESESESESESESESESESESESESESESESESESESESESESE 171

QY 237 F-DCLDKPIPMWRGLGACLAUTLIRSMKERVFCSTRTITTIQDGLSEYKQLGTQVLD 295
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 VKSAKD-----RYEEFISMETIKNAMS-----SNMIVVSNFPHLNKYSQ----- 214

QY 296 LKLVENHNFG--IHPWRVSSDLFPFASHAKYGYTLFPAQSHLEEVGKLNKYNHRLT- 352
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215 --KCKEAGRNPPYIEFLSALDQKLESADKA-----FIKSLDAANGRAFNAKQVRK 265

QY 353 -----MHPGOYTOIASPR-----EYVWDS-AIRDLAYHDEILSRM 386
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 266 NNRFQSDIDRYKDPGFMKPAELNEIPKPAKGAGODEVIVDGVATRGIVAPTEGLCK- 324

QY 387 KLNQLNKDAVLIHLGTFE--GKKTDLDRFRK-----NYQRLSDSVKARL 431
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 325 --PEEIT-PADIFKYLRAIFEARCKST-DRSEQIRLLEKLSITIAVTDYQRL--RVKVAL 378

QY 432 VLENDVSWSVQDILLPLCQ 450
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 379 LAVRFDINTSGOYMPIDQ 397
```

```
RESULT 7
YHML_YEAST
ID YHML_YEAST STANDARD; PRT; 607 AA.
AC Q04632; Q03641;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 69.8 KDA PROTEIN IN RPL6A-DAKI INTERGENIC REGION.
GN YML071C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE OF 1-113 FROM N.A.
RC STRAIN=S288C / AB972;
RA Bowman S., Churcher C., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 88-607 FROM N.A.
RC STRAIN=S288C / AB972;
RA Brown D., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z38114; CAAB6249.1; -
DR EMBL; Z46373; CAAB6507.1; -
DR SGD; S0004536; YML071C.
KW Hypothetical protein.
SQ SEQUENCE 607 AA; 69828 MW; 57E34E2DE36664B6 CRC64;
  4.7%; Score 145; DB 1; Length 607;
```

```
Best Local Similarity 21.6%; Pred. No. 0.57;
Matches 82; Conservative 73; Mismatches 160; Indels 64; Gaps 18;

QY 1 MURLLKRNTQIS--KRIVETILKQKAFKG-NHPCVPVSVCTIITYSRPHCLPDTL----- 50
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 258 LVKLLSTNYSVSLKKLQYLNSIPFDGKTNKSLLSV--FLAMRYKFTIDIASYPLDV 315

QY 51 ----KSLPMSSKTTLSMLPQ-----VNIGANSFSAET---PVDLKKENETELANISGPH 98
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 ESSNESLIEMMYKRKTEVLREHYVMSLVFLASFLYDNDLEIPPEELESVLINGTN 375

QY 99 KKSTSTSTRKARRSSKKKATDSVSKIDESAASYSTHLRSSRSKKPVYNSSSESES 158
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 EEKEIEKEKETKKEEYQKQDSVANN--EEDVTENKSIDVQVEVQCKVE--GEDDGAERKT 433

QY 159 EEOISKAT--KKYKQKEEEVEVEDEKSLKNSSSDSEFE--PVVPEOLETPISKRRRSR 215
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 434 ENIEENETVNTKEDKAEKBEVENTKONKAEKEEEENKVEVTPPEPSKSIDNKAEE--- 490

QY 216 SAKNLEKESMTNLDHAPREMFDCLDKPIPWGRGLGYACLTILRSMKERVFCSTRTCRIT 275
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 491 -----KEEIEINKVEVTPPEP-----SKKIRTSKRENKIPTNAVMLQFVDK-----C-IT 534

QY 276 TIORD---GLESVKQLGTQNVLDLJ----KLVEVHNHFGIHFMRVSSDLFPFASHAKYGY 328
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 535 YVLKDLTRGLNSIK-LSDSVCLQLVYCVSFRICDLNRNHHLEFLKKINDTSLF----- 585

QY 329 TLEFAQSHLEEVGKLANKY 347
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 586 TTEQLARADIKRAELASKY 604

RESULT 8
MLPL_YEAST
ID MLPL_YEAST STANDARD; PRT; 1875 AA.
AC Q02455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN-LIKE PROTEIN MLPL.
GN MLPL OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RC MEDLINE=93247549; PubMed=8483450;
RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Gen. Genet. 237:359-369(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MPLI genes and three
RT new open reading frames.";
RL Yeast 9:1349-1354(1993).
CC -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -!- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPLI".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```


[4] SEQUENCE OF 856-1940 FROM N.A.
 MEDLINE=8936648; PubMed=2771643;
 RA Karsch-Mizrachi I., Travis M., Blau H., Leinwand L.A.:
 RT "Expression and DNA sequence analysis of a human embryonic skeletal
 RL muscle myosin heavy chain gene."
 Nucleic Acids Res. 17:6167-6179(1989).
 CC -!- FUNCTION: MUSCLE CONTRACTION.
 CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
 CC -!- DEVELOPMENTAL STAGE: ABUNDANTLY PRESENT IN FETAL SKELETAL MUSCLE
 CC AND NOT PRESENT OR BARELY DETECTABLE IN HEART AND ADULT SKELETAL
 CC MUSCLE.
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMN) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X13988; CAA32167.1; -;
 DR EMBL; X13100; CAA31492.1; -;
 DR EMBL; X51593; CAA335942.1; -;
 DR EMBL; X15696; CAA33731.1; -;
 DR PIR; S04090; S04090.
 DR HSSP; P13538; 2MYS.
 DR MIN; 160720; -;
 DR InterPro; IPR000048; IQ.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR001609; myosin_head.
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSc; 1.
 DR PROSITE; PS50096; IQ; 1.
 DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
 KW Multigene family.
 FT DOMAIN 1 781 MYOSIN HEAD-LIKE.
 FT DOMAIN 782 811 IQ.
 FT DOMAIN 840 1933 COILED COIL (POTENTIAL).
 FT NP_BIND 179 186 ATP (POTENTIAL).
 FT DOMAIN 656 678 ACTIN-BINDING.
 FT DOMAIN 758 772 ACTIN-BINDING.
 FT MOD_RES 130 130 METHYLATION (TR1-) (POTENTIAL).
 FT MOD_RES 696 696 ALKYLATION (SH-1).
 FT MOD_RES 706 706 ALKYLATION (SH-2).
 FT CONFLICT 1331 1331 A -> G (IN REF. 3).
 FT CONFLICT 1391 1392 KK -> QE (IN REF. 1 AND 2).
 FT CONFLICT 1608 1609 SR -> RA (IN REF. 3).
 FT CONFLICT 1663 1664 RG -> QT (IN REF. 2).
 SQ SEQUENCE 1940 AA; 224035 MW; 43CA58C6A4BA1253 CRC64;

Query Match

4.4%; Score 138; DB 1; Length 1940;

Best Local Similarity 19.6%; Pred. No. 5.7;
 Matches 109; Conservative 97; Mismatches 206; Indels 144; Gaps 25;
 QY 84 KKNETELANISGPHKKSTSTSTSRKRARSKKK-ATDSVSDKIDSVASYDSS-THLRRS 141
 Db 941 KRKLEDECSEL---KDDDDLELTFLAKVEKEKHAENKVNKNTLESLGDELTIKLTRE 996
 QY 142 SRSKKPVNNSSESESEEQISKATKKVKQKEEVEEVDEKSLKNSSSSDEFEPVPE 201
 Db 997 KKALQEAHQALDQLAEDDKVNSLNKTKSKLEQQ-VEDL-ESSLEQE-----K 1043
 QY 202 QLETPISKRRSSSSAKNLEKSTMNIDHAPREMFDCLDKPIPWGRGLGYACLNILRS 261
 Db 1044 KLRVDLERNKRLKLEGLKLAQESILDLENDKQQ-----LDERL-----K 1082
 QY 262 MKRVRFCSTCRITTTIQRDGLD---SVKOLGTQNVLDLKLVEWNNHFIHFMVSSDLF 318
 Db 1083 KDFEYCCQLQSKVEDQDTLGLQFOKKIKELQAR-IEELEEEIAER----- 1127
 QY 319 PFASHAK-----YCYTLEFAQSHLEEVGKLANKNHRLTMHPGOYTQIA--SPREVVV 369
 Db 1128 --ATRAKTEKORSYARELEELSERLEEAGVTS-----TQIELNKKREAEF 1172
 QY 370 DSAIRDL---AYHDEILSKMKLNQNLNKDAVLIHLGCTFEGKKTLDLRFKNYORLSD 425
 Db 1173 LKLRRLDEATLOHEAMVATLR-----KHAHSVAVELG-----EQDNLQORVKOL-E 1219
 QY 426 SVKARLVLENDVMSVQ-----DLLPLCQELNPLVLDWHHHNIVPGTLREGSLDIM 478
 Db 1220 KKKSEFKLETDOLSSSMESVSMSKANLEKICRTLEDQL-----SEARGKNEEQRSLS 1272
 QY 479 PLIPTIREITWRKGLTQKOHYSADPTAISGMKRAHSDRVDFEPCDPTMDLMEAKE 538
 Db 1273 ELTOKSRLOTAGELSRLEEKESIVSLS-----RSKQAF-----TOOTBELKROLEE 1322
 QY 539 KEQAVFELCRRYELQNPCCLEIMGPEYDQTRGYPPGAEKRLTARKRS----- 589
 Db 1323 ENKAKNALAHALQSSSHDC--DLLREQYEEQEG---KAELOALRSKANSEVAQWRTKY 1376
 QY 590 -----RKEEVEDEK 599
 Db 1377 EYDATORTEEELEAKK 1392
 RESULT 12
 MSP1_PLAFK
 ID MSP1_PLAFK STANDARD; PRT: 1630 AA.
 AC P04932;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
 DE (PMMSA) (P190).
 GN MSP-1.
 OS Plasmodium falciparum (isolate K1 / Thailand).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5839;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=86136024; PubMed=3004972;
 RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certia U.,
 RA Stunnenberg H., Bujard H.;
 RT "Polymorphism of the precursor for the major surface antigens of
 RL Plasmodium falciparum merozoites: studies at the genetic level.";
 RN EMBO J. 4:3823-3829(1985).
 RN [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RA Pan W., Tolle R., Bujard H.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POTENTIAL).
 CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF

CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X03371; CAA27070.1; -
 DR PIR; A25120; SAZOK1.
 DR InterPro: IP0000561; EGF-like.
 DR Pfam; PF000008; EGF; 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1630 MEROZOITE SURFACE PROTEIN 1.
 FT DOMAIN 67 84 TRIPEPTIDE SG(TP) REPEAT.
 FT TRANSMEM 1614 1630 MEMBRANE ANCHOR.
 FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 755 755 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 759 759 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 774 774 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 835 835 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 911 911 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 955 955 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1049 1049 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1156 1156 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1436 1436 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1517 1517 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1630 AA; 187289 MW; ADBDEC3CE0A46322 CRC64;
 Query Match 4.4%; Score 136.5; DB 1; Length 1630;
 Best Local Similarity 19.5%; Pred. No. 5.5; Indels 217; Gaps 35;
 Matches 130; Conservativity 107; Mismatches 212; Indels 217; Gaps 35;
 Qy 17 FTILKOKAFKGNHPCVSVCTIYVS---REFHCLPDTLKLSPMSSKTTLSMLPQVNIGAN 73
 Db 176 FDLRLAKL--NDVCANDYCOIPENLIRANEL-DVLKLV----- 212
 Qy 74 SFSATPVDLKKENETELANISGPHKSTSTSTRKRARRSKKATDSVSKIDSVASYD 133
 Db 213 -FGYRKPLDNKDNVGMEDYI---KKN-----KKTIENINELIEESKKTID 255
 Qy 134 SSTHLRRSSRSKFPVN-----YNSSESESEBQISKATKKVKORFEERYVEEVEK-- 184
 Db 256 KKNATKEEKKLYQAYDLSYNNKOLE-EAHNLISVLEKRTDTLKNENIKELDKIN 314
 Qy 185 SLKN-----ESSDEPEVPVPEQLET-----PISKRRRSRSA 217
 Db 315 EIKNPPPPANSNGTPTNLLDNKKTEEHEKEIKETAKTIKENIDSLFTDPLEVEYLREKN 374
 Qy 218 KNLE-----KSTMLNLDHAPREMFDCIDKPIPWGRGLGYACLNTILISM----- 262
 Db 375 KNIDISAKVETKST-----EPNEYNGVYTP-----LSYNDINNALNELANSFGDLIN 422
 Qy 263 -----KERVFCSTRTCTITTDGLESVKOL---GTONVLDLILKVE-- 301
 Db 423 PFDYTKPSKNIYTDNERKKEIENIKIEKKIEKIESDKSYEDRSKSLNDITKEYEKL 482
 Qy 302 -----WNHFGIHFMRVSDLPFPAS---HAKYGYTLFPAQSHLEVEGKLANKYHRL 351
 Db 483 LNEIYDSKFNNNI-----DLTNFKMKMKRYSKYKVEKLTHTHTFASVENSCKHLEK 533
 Qy 352 THWPCQYQIASPREVVVDSAIRLAYHDETLRSM-----LNEQLNKDAVLIHLGGTF 406
 Db 534 LTKALKYMEDYSLRNIVVE---KELKYKNLISKIENEIETLVENIKKDEQL-----F 584
 Qy 407 EGKKTDLDRFRKNYQ--RLSDSVKAR-----LVLENDDVSNVQDILLPLCQLNPLVLWD 460

Db 585 E-KKITDKENPKDEKILEVSDIVKVOVKVLLMKNKIDELKKTKQLILK-----NVEL----- 634
 Qy 461 HHHNI-VPGTLREGS-----LD-----LMPLIPTIRETWTKGITQKHQHSSES 502
 Db 635 -KNIHIVNSVSKQENKQBPYYLVLVKEDIKLVKVPKVESLINEEKKNIKTEGO--SDN 691
 Qy 503 ADPTA---ISGMKRAHSDRVDFPPCPDPTMDLMTAEAKEKEQAVFELCRRRYELQNPCCPL 559
 Db 692 SEPTSEGEITQATTKPCQAGSALLEGD---SVQQAQAEQKQA-----QPPVPV 737
 Qy 560 ETMGPE 565
 Db 738 PV--PE 741
 RESULT 13
 MSP1_PLAFW
 ID MSP1_PLAFW STANDARD; PRT; 1639 AA.
 AC P04933;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
 DE (PMMSA) (P195).
 DE MSP-1.
 GN Plasmodium falciparum (isolate Wellcome).
 OS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RX NCBI_TaxID=5848;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86014355; PubMed=2995820;
 RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
 RA Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
 RA Freeman R.R.;
 RT "Primary structure of the precursor to the three major surface
 RT antigens of Plasmodium falciparum merozoites";
 RL Nature 317:270-273(1985).
 RN [2]
 RP REVISIONS.
 RA Holder A.A.;
 RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 (POTENTIAL).
 CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X02919; CAA26676.1; -
 DR PIR; A24594; A24594.
 DR InterPro: IP0000561; EGF-like.
 DR Pfam; PF000008; EGF; 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1639 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 268 268 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 764 764 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 768 768 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 783 783 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 844 844 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 920 920 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 964 964 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1058 1058 N-LINKED (GLCNAC. .) (POTENTIAL).

```

FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;

Query Match
Best Local Similarity 4.4%; Score 136.5; DB 1; Length 1639;
Matches 130; Conservative 107; Mismatches 212; Indels 217; Gaps 35;

QY 17 FTILKQAFKGNHPCVPSCTIYS---RFHCLPTLKLSPMSKTTLSMLPQVNTIGAN 73
Db 185 FOLLRAKL---NDVCANDYCOIPFNKLTRANEL-DVLKLV----- 221

QY 74 SFSAETPVDLKKENETELANISGPHKKSTSTRKRARSSKKKATDSVSDKIDESVASYD 133
Db 222 -FGYRKPLNDKNDGVKWDYI---KKN-----KKTLENINELIESKTTID 264

QY 134 SSTHLRRSRSKPNV-----YNSSESESEIOISATKKVKOKKEEEYVEEYDEK-- 184
Db 265 KKNKATKEEKKKLYQAOYLSYINKOLE-EAHNLISVLEKRIIDLTKKNENIKELLDKIN 323

QY 185 SLKN-----ESSDEREPVPVPEQLET-----PISKRERSRSA 217
Db 324 EIKNPPANSQNTPTNLLDKNNKIEHEKEKEIKETAKTIKFNIDSLFTDPLEYLYLREKN 383

QY 218 KNLE-----KESTMNLDHAPREMFDCLDPIPWGRGLGYACLTILRSM----- 262
Db 384 KNIDISAKVETKEST-----EPNEYPNGVTYP-----LSYNDINNALNELNFGDLIN 431

QY 263 -----KERVCSRCTRTITTQDGLSVKQL---GTQNVLDLKLVE-- 301
Db 432 PDYTKPSKNIYTDNERKKFTNEIKIEKKKIESDKSKYEDRSKSLNDITKEYEL 491

QY 302 -----WNHFGIHEMRVSSDLFPFAS--HAKYGYTLEPAQSHLEEVGKLANKYNHRL 351
Db 492 LNEIYDSKFNNI-----DLTFEKMGMKGRYKVEKLTHINTFASVENSXHNLEK 542

QY 352 TMHPQOYTQIASPREVVVDSDAIROLAYHDEILSRMK-----LNEOLNKDAVLIHLGGTF 406
Db 543 LTKALKYMEDYSLRNIVVE---KELKYKNLISKIENBIETLVENIKKDERQL-----F 593

QY 407 EGKKTETLDRFRKNYQ--RLSDSVKAR-----LVLENDVSVSVQDQLLPLCOELNIPVLWD 460
Db 594 E-KKITKDKNPDEKILEVSDIVKVQVKVLLMMNKIDELKKTQILK-----NVEL----- 643

QY 461 HHHNT-VPGTLREGS-----LD-----LMPLIPTIRETWTRKGIQKHYSSES 502
Db 644 -KHNIHVPNSYKQENKQEPYVYLLVILKKEIDKLKVPMPKVESLINEEKKNKITQEQ--SDN 700

QY 503 ADPTA---ISGMKRRASDRVDFDPPCDPTMDLMEAKEKEQAVFELCRRYELONPPCPL 559
Db 701 SEPSTGEITQATTKPGQOAGSALBGD---SVQAOAQBQKQA-----QPPVPV 746

QY 560 EIMGP 565
Db 747 PV--PE 750

RESULT 14
KELI_YEAST
ID KELI_YEAST STANDARD; PRT; 1164 AA.
AC P38853;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE KELCH REPEATS PROTEIN 1.
GN KELI OR YHR158C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;

```

```

[1]
SEQUENCE FROM N.A.
STRAIN-S288C / AB972;
MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
Lafaille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
Vaudin M.;
"Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
VIII."
Science 265:2077-2082(1994).
[2]
CHARACTERIZATION.
MEDLINE=99003296; PubMed=9786949;
RA Phillips J., Herskowitz I.;
"Identification of Kelp, a kelch domain-containing protein involved
in cell fusion and morphology in Saccharomyces cerevisiae.";
RL J. Cell Biol. 143:375-389(1998).
CC -!- FUNCTION: HAS A ROLE IN CELL MORPHOGENESIS AND CELL FUSION AND MAY
ANTAGONIZE THE PKC1 PATHWAY.
CC -!- SUBUNIT: INTERACTS WITH KEL2.
CC -!- SIMILARITY: CONTAINS 5 KELCH REPEATS.
CC -!- SIMILARITY: TO YEAST KEL2.
CC
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC
EMBL: U10397; AAC68991.1;
DR PIR: S46769; S46769.
DR SGD: S0001201; KELI.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF01344; Kelch; 3.
DR Repeat: Coiled coil.
KW REPEAT 139 186
FT REPEAT 253 307
FT REPEAT 308 357
FT REPEAT 359 409
FT REPEAT 411 460
FT REPEAT 477 531
FT DOMAIN 974 1163
FT COILED COIL (POTENTIAL).
FT COILED COIL (POTENTIAL).
SQ SEQUENCE 1164 AA; 131093 MW; 43D0FC570FD1D5E4D CRC64;

Query Match
Best Local Similarity 4.3%; Score 135; DB 1; Length 1164;
Matches 84; Conservative 64; Mismatches 147; Indels 94; Gaps 16;

QY 62 LSMLPOVNLGANSFSAETPVDLKKENETELANISGPHKKSTSTRKRARSSKKKATDSV 121
Db 616 LETLPSNEI-----KTP---QNCNIEIKHLPDADEKIDSTTTTDFDQEIINGDKLGTSM 665

QY 122 SDKIDE--SVASYDSSTHLRR--SSRSK---KPVNYSNSES--SESEEQTSKATKKVQKE 173
Db 666 S-KVEEDGNVADEDEIGVAQWASSPSKQDPKIKHYNESSELSONNTEIDKLS----- 717

QY 174 EEEYVEEVDKSLKNES--DEFPVVPQLETPISKRRSRSSAKNLEKSTMMNLDHA 232
Db 718 -----EPVDITTKSDTAGHDSANHVIDASDEKNVSPMGDVPDTDKNEEASVPINRD--A 770

QY 233 PREMFCDLKDPIPWGRGLGYACLTILRSMKERVCSRCTRTITTQDGLSVKQLGTON 292
Db 771 TTEVVD-----RALPEKL-----RSELQSLKELTHEK 797

QY 293 VLIDLKLVDNHNHFGTHFMRVSSDLFPFASHAKYGYTLEFAQ---SHLEEVGKLANKYNH 349

```

```
Db 798 ALEA-----GAHIKELETELWOLKSQKNSGTTKEIDELDSVRLOSKEILLEADNH 847
Qy 350 RUTMHPGQVQTASPREVVVDSAIRDLAYHDEILSRMKLNEOLNKDAVLIHLGGTFEGK 409
Db 848 SL-----EDKVNLEELVNSKFLDIENLEVI-----QFQNEKIKSLEPNYREK 893
Qy 410 KETLDRFRKNYQRLSDSVKARLVLENDV 438
Db 894 LEELOTEHENLSRENRLKNSKQHNEI 922

RESULT 15
POLR_ASCVP STANDARD; PRT: 2105 AA.
AC P36309;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: RNA REPLICASE (EC 2.7.7.48); HELICASE;
DE COAT PROTEIN].
OS Apple stem grooving virus (strain P-209) (ASGV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Capilliovirus.
OX NCBI_TaxID=36402;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93033164; PubMed=1413530;
RA Yoshikawa N., Sasaki E., Kato M., Takahashi T.;
RT "The nucleotide sequence of Apple stem grooving capilliovirus genome.";
RL Virology 191:98-105(1992).
CC 1- PTM: THE N-TERMINAL OF THE COAT PROTEIN IS BLOCKED.
CC 1- MISCELLANEOUS: THE COAT PROTEIN IS LOCATED IN THE CARBOXY-TERMINAL
CC REGION OF THIS POLYPROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D14995; BAA03639.1; -
CC PIR; A44059; A44059.
CC MEROPS; C35_001; -
KW RNA-directed RNA polymerase; Transferase; Polyprotein; ATP-binding;
FT NP_BIND 781 788 ATP (POTENTIAL).
FT DOMAIN 1364 1453 POLYMERASE ACTIVE SITE.
SQ SEQUENCE 2105 AA; 241240 MW; 331291A5FA137131 CRC64;

Query Match 4.3%; Score 134; DB 1; Length 2105;
Best Local Similarity 21.2%; Pred. No. 10;
Matches 113; Conservative 81; Mismatches 193; Indels 146; Gaps 28;

Qy 119 DSVSDKIDSVASVDSSTHLRRS-----SRSKKP-----VNVNSSSESESER 160
Db 491 NSGLKIDYSLVRLVARFRRLRDFLDIPSKGSPRKQFLFLESTGNIKSNPNAEKNSGES 550
Qy 161 QISKATKKVKOKEEVEVEVEVEVEVEVEVEVEVEVEVEVEVEVEVEVEVEVEVEVEVEVE 213
Db 551 -----IKIEGSAENDQHPHEVSHTSMTEDQG-FEGSIPVDLINCPEPEIKLPKRRK 603
Qy 214 R-----SAKNLEKESTWN--LDDHAPREMFDCLDKIPWGRGLGYACNLNLTIL----- 259
Db 604 NDCVFKAISAHLGDSQDLNLFVNEDISDELDCIEED---KG-LSHEMIEEVLITKGL 659
Qy 260 -----RSNKE-----RVFC-----SRTCRITTIORDG--LESVKOLGT 290
Db 660 SMVYTSDFKEMAVLNRYGVNGRMICYTIKGNHCELSKCEFIKLLKEGGQAQMSNENLNA 719
Qy 291 QNVLDLIKLVENHN-----FGIHF---MRVSSDLFPFASHAKY 326
```

Search completed: January 15, 2002, 14:07:05
Job time: 934 sec

```
Db 720 DSELDLGRFV---HNRDRAVKLAKSMARGTTGLLNEFDLEFCCKNMVTLSELEPFENFSSV 776
Qy 327 CYTLEPAQSHLEEVGKLANKNYHRLTMHPGOYTOIASPREVVVDSAIRDLAYHDEILSRM 386
Db 777 GLRLGFAGS-----GK-THKVLQWINTPVSVKRMFISPRMLA-----DEVEPOL 820
Qy 387 KLNE-QLNKDAVLIHLGGTF-EGKKTLDLFRKNYQRLSDSVKARLVLENDVSVSWQD 444
Db 821 KGTACQVHTWETALKKIDGTFMEVFVEIGLYPGYLTLLQMCFAFRKIVKQSENFLKKG 880
Qy 445 LLPLCQE-LNI-----PLVLDWHHHNVPGTLREGSLDLMLPLTIETWTRKGIQKH 498
Db 881 LLELSKTCNIRCFCDPLQL--RYSAEDTNLDKTHDIDLMIKTIKHKYLFQYRGQW 938
Qy 499 YSESAD-PTAI--SGMKRRAHSD----RVDFPPCDPTMDLMIEAKKEQAVF 544
Db 939 FOELVNMPTRVDESFKSKFFADISSVKTEDY-----GLILVAKREDKGVF 984
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:45:31 ; Search time 57.72 Seconds
(without alignments)
322.812 Million cell updates/sec

Title: US-09-724-296-2
Perfect score: 4349
Sequence: 1 MTKLPILGYNKIKGLVQPTR.....KRLTARKRRSRKEVEDEK 828

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues
Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
 - 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
 - 5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
 - 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1259.5	29.0	564	2	US-08-216-894-2
2	1259.5	29.0	564	4	US-09-115-746-2
3	1259.5	29.0	643	2	US-08-216-894-8
4	1259.5	29.0	643	4	US-09-115-746-8
5	1228.5	28.2	354	2	US-08-216-894-4
6	1228.5	28.2	354	4	US-09-115-746-4
7	1221.5	28.1	472	2	US-08-216-894-10
8	1221.5	28.1	472	4	US-09-115-746-10
9	1207.5	27.8	362	3	US-08-395-507-2
10	1206	27.7	426	3	US-08-737-248-4
11	1205.5	27.7	579	2	US-08-864-224-11
12	1204.5	27.7	515	4	US-08-974-549A-604
13	1203	27.7	514	4	US-08-974-549A-605
14	1201.5	27.6	352	1	US-08-395-507-1
15	1200.5	27.6	530	4	US-08-974-549A-603
16	1199.5	27.6	231	4	US-08-129-722A-2
17	1199.5	27.6	232	5	PCT-US94-09700-6
18	1199.5	27.6	337	5	PCT-US94-09700-11
19	1199.5	27.6	517	4	US-08-974-549A-606
20	1199.5	27.6	538	4	US-08-974-549A-602
21	1199.5	27.6	692	4	US-09-352-159-19
22	1199.5	27.6	692	4	US-09-352-168-19
23	1199.5	27.6	829	4	US-09-352-159-33
24	1199.5	27.6	829	4	US-09-352-168-33
25	1199.5	27.6	1196	4	US-09-352-159-31
26	1199.5	27.6	1196	4	US-09-352-168-31
27	1199.5	27.6	1205	4	US-09-352-159-29

Sequence 29, Appli
Sequence 1, Appli
Sequence 56, Appli
Sequence 6, Appli
Sequence 9, Appli
Sequence 4, Appli
Sequence 6, Appli
Sequence 3, Appli
Sequence 5, Appli
Sequence 11, Appli
Sequence 2, Appli
Sequence 26, Appli
Sequence 33, Appli
Patent No. 5427917
Sequence 4, Appli
Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-216-894-2
; Sequence 2, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Olsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 564 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-216-894-2

Query Match 29.0%; Score 1259.5; DB 2; Length 564;
Best Local Similarity 57.1%; Pred. No. 1.9e-94;
Matches 270; Conservative 55; Mismatches 103; Indels 45; Gaps 12;
QY 5 PTLGYWKIKGLVQPTLLLEYLEEKYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
Db 3 PTLGYWKIKGLVQPTLLLEYLEEKYERDEGDKWRNKKFELGLEFPNLPYYIDGD 62
QY 65 VKLTOSMAIRYIADKHNNLGGCPKRAFIEMLEGAVLDIRYGVSRVAYSKDFETLKVDF 124

```

Db 63 VKLTQSMALLIRYIADKHNMGLGCPKRAEISMLEGAVLDIRYGVSRAYSXDFETLKVD 122
QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHWTHDPDMYDALDVVLYMDPCLDAPFKLVCFK 184
Db 123 LSKLPEMLKMFEDRLCHKTYLNGDHWTHDPDMYDALDVVLYMDPCLDAPFKLVCFK 182
QY 185 IEAIPQIDKYLKSSKYIAMPLOGWATFGGDPKPSDHLVPRG--SMRLLLKRNQISK 242
Db 183 IEAIPQIDKYLKSSKYIAMPLOGWATFGGDPKPSD-LVPRGSPSLOQAENNTNSK 241
QY 243 RIVFTILKQAFKGNHPCVPSVCTIT-----YSRFHCLPD-----TLKSLLPMSSKT 289
Db 242 K-EMTKLREKVKAAEKELDAINRATKLEERNQAYKAAHKAEEBKAKTFQRLITFES 300
QY 290 -TSLMLPQVNI-----ANSFSAETPVDLKKENETELANISGPHKKTSTSTRKRARS 343
Db 301 INLKKRPDAVSNRDKKNSAKTID-EVEKQRAAEAAKAVETEKORAAEAT-KVAEAEK 358
QY 344 KATDSV-----SDKIDSVASYDSSHTLRRSSRKKPVNYSNSESSESEEQISKATKVK 399
Db 359 RKAEEAAKAVETEKORAAEATKVAEAEKQAAEAAKAV-----ETEKORAAEATKVA 410
QY 400 QKEEEYVEEVDKSLKNSSSDPEFVVPVPELETPIKRRRRSSSAKNLEKE 452
Db 411 EAEKQRAAEAMKVAEAEKQAA-----EATKVAEAEKQAAEATKVAEAE 455

```

RESULT 2

```

US-09-115-746-2
; Sequence 2, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYANOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115.746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 564 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-115-746-2

```

Query Match 29.0%; Score 1259.5; DB 4; Length 564;

```

Best Local Similarity 57.1%; Pred. No. 1.9e-94;
Matches 270; Conservative 55; Mismatches 103; Indels 45; Gaps 12;
QY 5 PILGYWKIKGLVQTRILLEYLBEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYIDGD 64
Db 3 PILGYWKIKGLVQTRILLEYLBEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYIDGD 62
QY 65 VKLTQSMALLIRYIADKHNMGLGCPKRAEISMLEGAVLDIRYGVSRAYSXDFETLKVD 124
Db 63 VKLTQSMALLIRYIADKHNMGLGCPKRAEISMLEGAVLDIRYGVSRAYSXDFETLKVD 122
QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHWTHDPDMYDALDVVLYMDPCLDAPFKLVCFK 184
Db 123 LSKLPEMLKMFEDRLCHKTYLNGDHWTHDPDMYDALDVVLYMDPCLDAPFKLVCFK 182
QY 185 IEAIPQIDKYLKSSKYIAMPLOGWATFGGDPKPSDHLVPRG--SMRLLLKRNQISK 242
Db 183 IEAIPQIDKYLKSSKYIAMPLOGWATFGGDPKPSD-LVPRGSPSLOQAENNTNSK 241
QY 243 RIVFTILKQAFKGNHPCVPSVCTIT-----YSRFHCLPD-----TLKSLLPMSSKT 289
Db 242 K-EMTKLREKVKAAEKELDAINRATKLEERNQAYKAAHKAEEBKAKTFQRLITFES 300
QY 290 -TSLMLPQVNI-----ANSFSAETPVDLKKENETELANISGPHKKTSTSTRKRARS 343
Db 301 INLKKRPDAVSNRDKKNSAKTID-EVEKQRAAEAAKAVETEKORAAEAT-KVAEAEK 358
QY 344 KATDSV-----SDKIDSVASYDSSHTLRRSSRKKPVNYSNSESSESEEQISKATKVK 399
Db 359 RKAEEAAKAVETEKORAAEATKVAEAEKQAAEAAKAV-----ETEKORAAEATKVA 410
QY 400 QKEEEYVEEVDKSLKNSSSDPEFVVPVPELETPIKRRRRSSSAKNLEKE 452
Db 411 EAEKQRAAEAMKVAEAEKQAA-----EATKVAEAEKQAAEATKVAEAE 455

```

RESULT 3

```

US-08-216-894-8
; Sequence 8, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYANOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```


MOLECULE TYPE: protein
US-08-216-894-8

Query Match 29.0%; Score 1259.5; DB 2; Length 643;
Best Local Similarity 57.1%; Pred. No. 2.3e-94;
Matches 270; Conservative 55; Mismatches 103; Indels 45; Gaps 12;

```

QY 5 PILGYWKIKGLVQPTRLLEYLEEKYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
DB 3 PILGYWKIKGLVQPTRLLEYLEEKYERDEGDKWRNKKFELGLEFPNLPYYIDGD 62
QY 65 VKLTQSMAIRIYADKHNMLGGCPKERAISMLEGAVLDIRYGVSR IAYSKDFETLKVD 124
DB 63 VKLTQSMAIRIYADKHNMLGGCPKERAISMLEGAVLDIRYGVSR IAYSKDFETLKVD 122
QY 125 LSKLPEMLKMPEDRLCHKTYLNGDHTVHPDFMLYDALDVVLYMDPMDCLDAFPKLVCFKKR 184
DB 123 LSKLPEMLKMPEDRLCHKTYLNGDHTVHPDFMLYDALDVVLYMDPMDCLDAFPKLVCFKKR 182
QY 185 IEAIPQIDKYLKSSKYIAMPLOGQWATFGGDHPPKSDHLPVPG--SMLRLKRNIOISK 242
DB 183 IEAIPQIDKYLKSSKYIAMPLOGQWATFGGDHPPKSDHLPVPG--SMLRLKRNIOISK 241
QY 243 RIVFTTLKOKAFKGNHPCVPSVCTIT-----YSRFHCLPD---TLKSLLPMSST 289
DB 242 K-EMTKLREKVKKAEKLDINRATKLEERNOAYKAAHKAEEKAKTFFORLITFESN 300
QY 290 -TSLMLPQVNI-----ANSFSAETPVDLKKNETELANISGPHKKSTSTSTRKRARSSK 343
DB 301 INLKRPNDAVSNRDKKNSSETAKTD-EVEKQRAAEAKAVETEKQRAAEAT-KVAEAEK 358
QY 344 KKATDSV-----SDKIDESVASYDSSTHLRRSRKPKPVNYSSESESEEQISKATKKVK 399
DB 359 RKAEEAAKAVETEKQRAAEATKVAEAKOKAAEAAKAV-----ETEKQRAAEATKVA 410
QY 400 QKEEEYVEEVDKSLKNESSEDFEPVPEQLETPISKRRSRSSAKNLEKE 452
DB 411 EAEKQRAAEAMKVAEAKOKAA-----EATKVAEAKOKAAEATKVAEAE 455

```

RESULT 4

```

US-09-115-746-8
: Sequence 8, Application us/09115746
: Patent No. 6228601
: GENERAL INFORMATION:
: APPLICANT: Kirchhoff, Louis V.
: APPLICANT: Otsu, Keiko
: TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W., Suite 500
: CITY: Washington, D.C.
: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/115,746
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/216,894
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 85326/102/DRLO
: TELECOMMUNICATION INFORMATION:

```

TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 643 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-115-746-8

Query Match 29.0%; Score 1259.5; DB 4; Length 643;
Best Local Similarity 57.1%; Pred. No. 2.3e-94;
Matches 270; Conservative 55; Mismatches 103; Indels 45; Gaps 12;

```

QY 5 PILGYWKIKGLVQPTRLLEYLEEKYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
DB 3 PILGYWKIKGLVQPTRLLEYLEEKYERDEGDKWRNKKFELGLEFPNLPYYIDGD 62
QY 65 VKLTQSMAIRIYADKHNMLGGCPKERAISMLEGAVLDIRYGVSR IAYSKDFETLKVD 124
DB 63 VKLTQSMAIRIYADKHNMLGGCPKERAISMLEGAVLDIRYGVSR IAYSKDFETLKVD 122
QY 125 LSKLPEMLKMPEDRLCHKTYLNGDHTVHPDFMLYDALDVVLYMDPMDCLDAFPKLVCFKKR 184
DB 123 LSKLPEMLKMPEDRLCHKTYLNGDHTVHPDFMLYDALDVVLYMDPMDCLDAFPKLVCFKKR 182
QY 185 IEAIPQIDKYLKSSKYIAMPLOGQWATFGGDHPPKSDHLPVPG--SMLRLKRNIOISK 242
DB 183 IEAIPQIDKYLKSSKYIAMPLOGQWATFGGDHPPKSDHLPVPG--SMLRLKRNIOISK 241
QY 243 RIVFTTLKOKAFKGNHPCVPSVCTIT-----YSRFHCLPD---TLKSLLPMSST 289
DB 242 K-EMTKLREKVKKAEKLDINRATKLEERNOAYKAAHKAEEKAKTFFORLITFESN 300
QY 290 -TSLMLPQVNI-----ANSFSAETPVDLKKNETELANISGPHKKSTSTSTRKRARSSK 343
DB 301 INLKRPNDAVSNRDKKNSSETAKTD-EVEKQRAAEAKAVETEKQRAAEAT-KVAEAEK 358
QY 344 KKATDSV-----SDKIDESVASYDSSTHLRRSRKPKPVNYSSESESEEQISKATKKVK 399
DB 359 RKAEEAAKAVETEKQRAAEATKVAEAKOKAAEAAKAV-----ETEKQRAAEATKVA 410
QY 400 QKEEEYVEEVDKSLKNESSEDFEPVPEQLETPISKRRSRSSAKNLEKE 452
DB 411 EAEKQRAAEAMKVAEAKOKAA-----EATKVAEAKOKAAEATKVAEAE 455

```

RESULT 5

```

US-08-216-894-4
: Sequence 4, Application US/08216894
: Patent No. 5876734
: GENERAL INFORMATION:
: APPLICANT: Kirchhoff, Louis V.
: APPLICANT: Otsu, Keiko
: TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W., Suite 500
: CITY: Washington, D.C.
: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/216,894
: FILING DATE: 24-MAR-1994

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-216-894-4

```

```

Query Match      28.2%; Score 1228.5; DB 2; Length 354;
Best Local Similarity 69.3%; Pred. No. 3.1e-92;
Matches 250; Conservative 24; Mismatches 60; Indels 27; Gaps 7;

```

```

QY 5 PILGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
   |||||
DB 3 PILGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGLEFPNLPYYIDGD 62
   |||||
QY 65 VKLTOSMAIIRIADKHNLGCGPKERAELISMLEGAVLDIRYGVSRAYSKDFTLKVDF 124
   |||||
DB 63 VKLTOSMAIIRIADKHNLGCGPKERAELISMLEGAVLDIRYGVSRAYSKDFTLKVDF 122
   |||||
QY 125 LSKLPEMLKMFEDRLCHTKTYLNGDHVTHPDMLYDALDVLYMDPCLDAPPKLVCFKKR 184
   |||||
DB 123 LSKLPEMLKMFEDRLCHTKTYLNGDHVTHPDMLYDALDVLYMDPCLDAPPKLVCFKKR 182
   |||||
QY 185 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDHPPKSDHLPVPG--SMRLRLKRNIOISK 242
   |||||
DB 183 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDHPPKSD--LVPRSPSOLQOAEENITNSK 241
   |||||
QY 243 RIVFTILKOKAFKGNHPCVPSVCTIT-----YSRFHCLPD-----TLKSLPMSKKT 289
   |||||
DB 242 K-EMTKLEKVKKAKEKLDINRATKLEERNQAYKAAHKAEEKAKTFORLITFESE- 299
   |||||
QY 290 TSLMLPQVNIAGNSFAETPVDLKKENETELANISGPHKSTSTSTRKARSKKKATDS 349
   |||||
DB 300 -----NINLKKRPNDVSNRDKKNSKSETAKTD-----EVEKQRAAAKAVETEKORAGEF 350
   |||||
QY 350 V 350
DB 351 I 351

```

```

RESULT 6
US-09-115-746-4
; Sequence 4, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115.746

```

```

; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-115-746-4

```

```

Query Match      28.2%; Score 1228.5; DB 4; Length 354;
Best Local Similarity 69.3%; Pred. No. 3.1e-92;
Matches 250; Conservative 24; Mismatches 60; Indels 27; Gaps 7;

```

```

QY 5 PILGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
   |||||
DB 3 PILGYWKIKGLVQPTRLLEYLEEKEEHLERDEGDKWRNKKFELGLEFPNLPYYIDGD 62
   |||||
QY 65 VKLTOSMAIIRIADKHNLGCGPKERAELISMLEGAVLDIRYGVSRAYSKDFTLKVDF 124
   |||||
DB 63 VKLTOSMAIIRIADKHNLGCGPKERAELISMLEGAVLDIRYGVSRAYSKDFTLKVDF 122
   |||||
QY 125 LSKLPEMLKMFEDRLCHTKTYLNGDHVTHPDMLYDALDVLYMDPCLDAPPKLVCFKKR 184
   |||||
DB 123 LSKLPEMLKMFEDRLCHTKTYLNGDHVTHPDMLYDALDVLYMDPCLDAPPKLVCFKKR 182
   |||||
QY 185 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDHPPKSDHLPVPG--SMRLRLKRNIOISK 242
   |||||
DB 183 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDHPPKSD--LVPRSPSOLQOAEENITNSK 241
   |||||
QY 243 RIVFTILKOKAFKGNHPCVPSVCTIT-----YSRFHCLPD-----TLKSLPMSKKT 289
   |||||
DB 242 K-EMTKLEKVKKAKEKLDINRATKLEERNQAYKAAHKAEEKAKTFORLITFESE- 299
   |||||
QY 290 TSLMLPQVNIAGNSFAETPVDLKKENETELANISGPHKSTSTSTRKARSKKKATDS 349
   |||||
DB 300 -----NINLKKRPNDVSNRDKKNSKSETAKTD-----EVEKQRAAAKAVETEKORAGEF 350
   |||||
QY 350 V 350
DB 351 I 351

```

```

RESULT 7
US-08-216-894-10
; Sequence 10, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-216-894-10

```

```

Query Match 28.1%; Score 1221.5; DB 2; Length 472;
Best Local Similarity 56.7%; Pred. No. 1.8e-91;
Matches 255; Conservative 54; Mismatches 110; Indels 31; Gaps 9;

QY 5 PILGYWKIKGLVQTRLLLEYLEEYEHLYERDEGDKWRNKKFELGLFPNLPYYIDGD 64
DB 3 PILGYWKIKGLVQTRLLLEYLEEYEHLYERDEGDKWRNKKFELGLFPNLPYYIDGD 62
QY 65 VKLTQSMAIRIYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYASKDFETLKVDF 124
DB 63 VKLTQSMAIRIYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYASKDFETLKVDF 122
QY 125 LSKLPEMLKMPEDRLCHKTLYLNGDHVTHPDFMLYDALDVVLYMDPMDCLDAFPKLVCFKKR 184
DB 123 LSKLPEMLKMPEDRLCHKTLYLNGDHVTHPDFMLYDALDVVLYMDPMDCLDAFPKLVCFKKR 182
QY 185 IEAIPQIDKYLLKSSKYIAMPLOGHQATFGGDDHPPKSDHLVPRGSMRLRLKRNIIQISKRI 244
DB 183 IEAIPQIDKYLLKSSKYIAMPLOGHQATFGGDDHPPKSDHLVPRGSMRLRLKRNIIQISKRI 237
QY 245 VFTILKQAFKGNHPCVPSVCTITYSRFHCLPDTLKLSPMSKTTLSMLPQVNIANSF 304
DB 238 A-EAEKQAAEAT-----KVAEAEKQAAEATK--VAEAEKQAAEATKVAEAEKQOR 286
QY 305 SAETP--VDLKENETELANISGPHKSTSTSTRKRARRSKKATDSVSKIDESVASYD 362
DB 287 AAEATKVAEAEKQAAEATKVADEKQAAEAT--KVAEAEKQAAEAT--KVAE--AEKQ 341
QY 363 SSTHLRRSSRSKPKVNYNSSESESEBQISKATKKYKQKEEEYVEEVEDEKSLKNSSSD 422
DB 342 KAAEATKVAEAEKQAAEATKVAEAEKQAAEATKVAEAEKQAAEATKVAEAEKQKAA- 400
QY 423 EFEPVVPQLETPISKRRSSRSKAKNLEKE 452
DB 401 -----EATKVAEAEKQAAEATKVAEAE 423

```

```

RESULT 8
; Sequence 10, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.

```

```

; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-115-746-10

```

```

Query Match 28.1%; Score 1221.5; DB 4; Length 472;
Best Local Similarity 56.7%; Pred. No. 1.8e-91;
Matches 255; Conservative 54; Mismatches 110; Indels 31; Gaps 9;

QY 5 PILGYWKIKGLVQTRLLLEYLEEYEHLYERDEGDKWRNKKFELGLFPNLPYYIDGD 64
DB 3 PILGYWKIKGLVQTRLLLEYLEEYEHLYERDEGDKWRNKKFELGLFPNLPYYIDGD 62
QY 65 VKLTQSMAIRIYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYASKDFETLKVDF 124
DB 63 VKLTQSMAIRIYIADKHNMLGGCPKRAEISMLEGAVLDIRYGVSRISYASKDFETLKVDF 122
QY 125 LSKLPEMLKMPEDRLCHKTLYLNGDHVTHPDFMLYDALDVVLYMDPMDCLDAFPKLVCFKKR 184
DB 123 LSKLPEMLKMPEDRLCHKTLYLNGDHVTHPDFMLYDALDVVLYMDPMDCLDAFPKLVCFKKR 182
QY 185 IEAIPQIDKYLLKSSKYIAMPLOGHQATFGGDDHPPKSDHLVPRGSMRLRLKRNIIQISKRI 244
DB 183 IEAIPQIDKYLLKSSKYIAMPLOGHQATFGGDDHPPKSDHLVPRGSMRLRLKRNIIQISKRI 237
QY 245 VFTILKQAFKGNHPCVPSVCTITYSRFHCLPDTLKLSPMSKTTLSMLPQVNIANSF 304
DB 238 A-EAEKQAAEAT-----KVAEAEKQAAEATK--VAEAEKQAAEATKVAEAEKQOR 286
QY 305 SAETP--VDLKENETELANISGPHKSTSTSTRKRARRSKKATDSVSKIDESVASYD 362
DB 287 AAEATKVAEAEKQAAEATKVADEKQAAEAT--KVAEAEKQAAEAT--KVAE--AEKQ 341
QY 363 SSTHLRRSSRSKPKVNYNSSESESEBQISKATKKYKQKEEEYVEEVEDEKSLKNSSSD 422
DB 342 KAAEATKVAEAEKQAAEATKVAEAEKQAAEATKVAEAEKQAAEATKVAEAEKQKAA- 400
QY 423 EFEPVVPQLETPISKRRSSRSKAKNLEKE 452
DB 401 -----EATKVAEAEKQAAEATKVAEAE 423

```

```

RESULT 9
; US-08-395-507-2
; Sequence 2, Application US/08395507
; Patent No. 5578456
; GENERAL INFORMATION:
; APPLICANT: Fujimura, Katsuya

```



```

; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..515
; OTHER INFORMATION: /note= "fusion protein composed of
; OTHER INFORMATION: glutathione-S-transferase sequence,
; OTHER INFORMATION: thrombin cleavage sequence, recognition
; OTHER INFORMATION: sequence for heart muscle protein
; OTHER INFORMATION: kinase, residues introduced by cloning
; OTHER INFORMATION: and hprt protein fragment"
; US-08-974-549A-604

Query Match 27.7%; Score 1204.5; DB 4; Length 515;
Best Local Similarity 69.7%; Pred. No. 5e-90;
Matches 241; Conservative 17; Mismatches 33; Indels 55; Gaps 5;

QY 5 PILGWKIKGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGFPNLPYYIDG 64
Db 3 PILGWKIKGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGFPNLPYYIDG 62
QY 65 VKLTQSMATIRYIADKHNLGGCPKRAEISMLEGAVLDIRYGVSRVIAYSKDFETLKVD 124
Db 63 VKLTQSMATIRYIADKHNLGGCPKRAEISMLEGAVLDIRYGVSRVIAYSKDFETLKVD 122
QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHTVTHPDFMLYDALDVLVYMDPCLDAFFPKLVCFKR 184
Db 123 LSKLPEMLKMFEDRLCHKTYLNGDHTVTHPDFMLYDALDVLVYMDPCLDAFFPKLVCFKR 182
QY 185 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDDHPPKSDHLPVGRS-----M 230
Db 183 IEAIPQIDKYLKSSKYIAWPLQGWQATFGGDDHPPKSD-LVPRGSRASVGSYTKMSYV 241
QY 231 LRLKRNQIS-----KRIVETILKOKAFKGNHPCV 261
Db 242 VELURSFYVYTTFOKNLFFRPSVWSKLSQISIGIRQLKRVQLRELSEAEVRQHREAR 301
QY 262 PSYCTIYSRFLCPDLTKSLPLMSKSTLTLMLPVNIGANSFSAE 307
Db 302 PALLT-SRLRFIPKDLRPVNM-----DYVVGARTFERE 336

RESULT 13
US-08-974-549A-605
; Sequence 605, Application US/08974549A
; Patent No. 6166178
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin B.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,549A

```

```

; FILING DATE: 19-NOV-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/911,312
; FILING DATE: 14-AUG-1997
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/912,951
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/915,503
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/17618
; FILING DATE: 01-OCT-1997
; APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/17885
; FILING DATE: 01-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph Red
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002610US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 605:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 514 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..515
; OTHER INFORMATION: /note= "fusion protein composed of
; OTHER INFORMATION: glutathione-S-transferase sequence,
; OTHER INFORMATION: thrombin cleavage sequence, recognition
; OTHER INFORMATION: sequence for heart muscle protein
; OTHER INFORMATION: kinase, residues introduced by cloning
; OTHER INFORMATION: and hprt protein fragment"
; US-08-974-549A-605

```

```

Query Match 27.7%; Score 1203; DB 4; Length 514;
Best Local Similarity 71.6%; Pred. No. 6.6e-90;
Matches 244; Conservative 13; Mismatches 36; Indels 48; Gaps 7;

QY 5 PILGWKIKGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGFPNLPYYIDG 64
Db 3 PILGWKIKGLVQPTRLLEYLEEYEEHLYERDEGDKWRNKKFELGFPNLPYYIDG 62
QY 65 VKLTQSMATIRYIADKHNLGGCPKRAEISMLEGAVLDIRYGVSRVIAYSKDFETLKVD 124
Db 63 VKLTQSMATIRYIADKHNLGGCPKRAEISMLEGAVLDIRYGVSRVIAYSKDFETLKVD 122
QY 125 LSKLPEMLKMFEDRLCHKTYLNGDHTVTHPDFMLYDALDVLVYMDPCLDAFFPKLVCFKR 184
Db 123 LSKLPEMLKMFEDRLCHKTYLNGDHTVTHPDFMLYDALDVLVYMDPCLDAFFPKLVCFKR 182

```

QY 185 IEAIPQIDKYLKSSKYIAMPLOGWQATCGGDPKPSDHLVPRGSMRLKRNIOISKRI 244
DB 183 IEAIPQIDKYLKSSKYIAMPLOGWQATCGGDPKPSD-LVPRGS-----RRAS 230
QY 245 VFTLKOKAFKG-----NHPKV-----PSV-----CTITYSRF-HCL---- 275
DB 231 VGSVTKATSEALSGSTRHSFVSGRQHAGPPSRSRPRWDTPCPPVYATKHFLYSS 290
QY 276 -----POTLSLPLMSKTTLSMLPQVNICANSFSAETP 309
DB 291 GQKQELRPSFLLSLRPSLTGARRLVETITFLGSRPWPMPGTP 331

RESULT 14

US-08-395-507-1
; Sequence 1, Application US/08395507
; Patent No. 5578456

GENERAL INFORMATION:

APPLICANT: Fujimura, Katsuya
APPLICANT: Ueno, Eichi
APPLICANT: Fujii, No. 5578456uyuki
APPLICANT: Okada, Masahisa
TITLE OF INVENTION: Anti-treponema Pallidum Antibody
TITLE OF INVENTION: Immunoassay
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pollock, Vande Sande & Priddy
STREET: 1990 M Street, NW, Suite 800
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/395,507
FILING DATE: 27-FEB-1995
CLASSIFICATION: 436

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP Hei-6-54672
FILING DATE: 28-FEB-1995

ATTORNEY/AGENT INFORMATION:

NAME: Amernick, Burton A.
REGISTRATION NUMBER: 24,852
REFERENCE/DOCKET NUMBER: 1581/007
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 331-7111
TELEFAX: (202) 293-6229

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 352 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: rabbit
FEATURE:
NAME/KEY: Protein
LOCATION: 1..352
OTHER INFORMATION: /note= "G15 Antigen"

US-08-395-507-1

Query Match 27.6%; Score 1201.5; DB 1: Length 352;
Best Local Similarity 99.1%; Pred. No. 5e-90;
Matches 225; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 5 PILGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYIDGD 64
DB 3 PILGYWKIKGLVQPTRLLLEYLEEKYEELHYERDEGDKWRNKKFELGLEFPNLPYYIDGD 62
QY 65 VKLTOSMAITRYIADKNIMLGGCPKRAEISMLEGAVLDIRYGVSRITAYSKDFETLKVDF 124
DB 63 VKLTOSMAITRYIADKNIMLGGCPKRAEISMLEGAVLDIRYGVSRITAYSKDFETLKVDF 122
QY 125 LSKLPEMLKMFEDRLCHTKYTLNGDHTVDPDMLYDALDVLVYMDPMLCDLAPFKLVCFKKR 184
DB 123 LSKLPEMLKMFEDRLCHTKYTLNGDHTVDPDMLYDALDVLVYMDPMLCDLAPFKLVCFKKR 182
QY 185 IEAIPQIDKYLKSSKYIAMPLOGWQATCGGDPKPSDHLVPRGSMRL 231
DB 183 IEAIPQIDKYLKSSKYIAMPLOGWQATCGGDPKPSD-LVPRGSEL 228

RESULT 15

US-08-974-549A-603
; Sequence 603, Application US/08974549A
; Patent No. 6166178

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin B.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
PRIORITY APPLICATION DATA: WO PCT/US97/17885
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 603:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..530
OTHER INFORMATION: /note= "fusion protein composed of
OTHER INFORMATION: glutathione-S-transferase sequence,
OTHER INFORMATION: thrombin cleavage sequence, recognition
OTHER INFORMATION: sequence for heart muscle protein
OTHER INFORMATION: kinase, residues introduced by cloning,
OTHER INFORMATION: eight consecutive His residues and hTbT
OTHER INFORMATION: protein fragment"
US-08-974-549A-603

Search completed: January 15, 2002, 13:50:57
Job time: 326 sec

Query Match 27.6%; Score 1200.5; DB 4; Length 530;
Best Local Similarity 50.4%; Pred. No. 1.1e-89;
Matches 263; Conservative 49; Mismatches 91; Indels 119; Gaps 11;
QY 5 PILGYWKIKGLVQPTLLLEYLPEEYEEHLYERDECDKWRNKKFELGLEFPNLPYYIDGD 64
Db |||||
QY 3 PILGYWKIKGLVQPTLLLEYLPEEYEEHLYERDECDKWRNKKFELGLEFPNLPYYIDGD 62
Db |||||
QY 65 VKLTQSMATIRYIADKHNHMGCCPKERAESMLEGAVLDIRYGVSRISYKDFETLKVDF 124
Db |||||
QY 63 VKLTQSMATIRYIADKHNHMGCCPKERAESMLEGAVLDIRYGVSRISYKDFETLKVDF 122
Db |||||
QY 125 LSKLPELAKMFEDRLCHKTYLNGDHVTHPDEFMLYDALDVLVMDPCLDAPPKLVCFKKR 184
Db |||||
QY 123 LSKLPELAKMFEDRLCHKTYLNGDHVTHPDEFMLYDALDVLVMDPCLDAPPKLVCFKKR 182
Db |||||
QY 185 IEAIPQIDKYLKSSKYIAMPLOGWQATGGGDPKSDHLPVPGS----- 229
Db |||||
QY 183 IEAIPQIDKYLKSSKYIAMPLOGWQATGGGDPKSD- LYPGRSRRASVGSVHHHHHHH 241
Db |||||
QY 230 -----MLRLKRNTOIS-----KRIVFTILK 250
Db ::|||
QY 242 HGSVTKMSVYVVELLSRFYVTTETTFQKNRLFFYRPSVMSKLSQSIGIRQHLKRVQLRELS 301
Db ::|||
QY 251 QKAFKGNHPCVSVCTITYSRFCHLPDTLLKSLPMSSKTTLSMLPQVNVIGANSFSAETPV 310
Db ::|||
QY 302 EAEVQHRERAPALT-SKLRFIPKPDGLRPIVM-----DYVVGARTFRKKRA 350
Db ::|||
QY 311 D-----LKKENE-----TELANISGPHKSTSTTKRARSKKK----- 345
Db ::|||
QY 351 ERLTSRVKALFVNLVYERARRPGLLGASVLGDDITHRAWRTFVLVRVRAQDPPELYFVKV 410
Db ::|||
QY 346 ----ATDSV-SDKIDESVASY-----DSSTHLRRSRKSKKPNYNSSS 383
Db ::|||
QY 411 DVTGAYDIPQDRUTEVITASIKPONTYCVRRYAVVOKAAHGVHKKAFK-----HVST 464
Db ::|||
QY 384 ESESEQISKATKKYKQKEEYEEVDEKSLKNSSSDEFE 425
Db ::|||
QY 465 LFDLPYMRQFVAHLQETSPLRDAVWIEQSSSLNEASSGLFD 506
Db ::|||

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	360	18.2	320	1	YWJD_BACSU	P45864	bacillus su
2	111.5	5.6	517	1	VATB_YEAST	P16140	saccharomyc
3	110	5.6	463	1	VATB_DESSY	O06505	desulfuroco
4	108.5	5.5	492	1	VATB_ARATH	P11574	arabidopsis
5	108	5.5	500	1	VATB_CYAC	P84813	cyanidium c
6	104.5	5.3	470	1	VATB_ARCFU	O29100	archaeoglob
7	104	5.3	463	1	VATB_THESI	O32467	thermococcu
8	104	5.3	490	1	VATB_DROME	P31409	thermophilu
9	103.5	5.2	2230	1	GOGA_HUMAN	Q13439	homo sapien
10	103	5.2	494	1	VATB_HELVI	P13410	heliolithis v
11	103	5.2	494	1	VATB_MANSE	P14011	manduca sex
12	102	5.2	513	1	VAB1_BOVIN	P14007	bos taurus
13	101.5	5.1	453	1	VATB_CHICK	P49712	gallus gall
14	101.5	5.1	465	1	VATB_PYRAB	O94XU8	pyrococcus
15	101	5.1	491	1	VATB_CAEEL	Q19628	caenorhabdi
16	101	5.1	800	1	CYAA_BACAN	P40136	bacillus an
17	100.5	5.1	497	1	PROD_MOUSE	Q9WU79	mus musculu
18	100.5	5.1	511	1	VAB2_HUMAN	P21281	homo sapien
19	100.5	5.1	511	1	VAB2_MOUSE	P50517	mus musculu
20	100	5.1	373	1	BIOF_AQUAE	O66875	aquifex aeo
21	100	5.1	513	1	VAB1_HUMAN	P15133	homo sapien
22	99.5	5.0	273	1	Y133_METJA	Q57597	methanococc
23	98.5	5.0	488	1	VAB1_GOSHI	Q43432	gossypium h
24	97.5	4.9	386	1	VAB2_GOSHI	Q43433	gossypium h
25	97.5	4.9	511	1	VAB2_BOVIN	P31408	bos taurus
26	96.5	4.9	465	1	VATB_PYRHO	O57729	pyrococcus
27	96	4.9	511	1	VATB_CANTR	P22550	candida tro
28	95	4.8	365	1	GCPE_PROST	P72241	providencia
29	95	4.8	4036	1	RRPL_DUGBV	O66431	dugbe virus
30	94.5	4.8	492	1	VAB2_ACEAT	Q38681	acetabulari
31	93.5	4.7	492	1	VAB1_ACEAT	Q38680	acetabulari
32	93.5	4.7	648	1	NCPI_MOUSE	P28660	mus musculu
33	93.5	4.7	1128	1	NCPI_RAT	P55161	rattus norv

[illegible]

Qy 280

```

Db 257 RSHADYVDANFLPLLERFRQWGTNIDFMIEAKQKQKALLRL 298

RESULT 2
VATB_YEAST
ID VATB_YEAST STANDARD; PRT; 517 AA.
AC P16140; P32123;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE VACUOLAR ATP SYNTHASE SUBUNIT B (EC 3.6.1.34) (V-ATPASE B SUBUNIT)
DE (VACUOLAR PROTON PUMP B SUBUNIT) (V-ATPASE 57 KDA SUBUNIT).
GN VMA2 OR VAT2 OR YBR137C OR YBR1002.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89109197; PubMed=2521486;
RA Nelson H., Mandiyan S., Nelson N.;
RA "A conserved gene encoding the 57-kDa subunit of the yeast vacuolar
RT H+-ATPase."
RL J. Biol. Chem. 264:1775-1778(1989).
RN [2]
RP REVISIONS.
RA Nelson H., Mandiyan S., Nelson N.;
RL J. Biol. Chem. 264:5313-5313(1989).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C;
RA Becam A.-M., Cullin C., Grzybowski E., Lacroix F., Nasr F.,
RA Ozier-kalogeropoulos O., Palucha A., Slonimski P.P., Zagulska M.,
RA Herbert C.J.;
RT "The sequence of 29.7 kb from the right arm of chromosome II reveals
RT 13 complete open reading frames, of which ten correspond to new
RL genes."
RL Yeast 10:S1-S11(1994).
RN [4]
RP SEQUENCE OF 103-517 FROM N.A.
RX MEDLINE=92156101; PubMed=1371275;
RA Puopolo K., Kumamoto T., Adachi I., Mager R., Forgac M.;
RT "Differential expression of the 'B' subunit of the vacuolar
RT H(+)-ATPase in bovine tissues."
RL J. Biol. Chem. 267:3696-3706(1992).
CC -!- FUNCTION: NONCATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF
CC VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS. IT IS AN
CC ELECTROGENIC PROTON PUMP THAT GENERATES A PROTON MOTIVE FORCE OF
CC 180 MV, INSIDE POSITIVE AND ACIDIC, IN THE VACUOLAR MEMBRANE
CC VESICLES.
CC -!- SUBUNIT: V-ATPASE IS A HETEROMULTIMERIC ENZYME COMPOSED OF A
CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,
CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC -!- CAUTION: REF.4 SEQUENCE WAS INCORRECTLY THOUGHT TO ORIGINATE FROM
CC BOVINE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J04450; AAA6890.1; -
CC EMBL: X75891; CAAS3486.1; -
CC EMBL: Z35996; CAAS5084.1; -
CC EMBL: M8130; AAA30389.1; -
CC PIR: S45996; S45996.

PIR: B42254; B42254.
PIR: S46570; S46570.
DR YRPD; 8560; -.
DR SGD; S0000331; VMA2.
DR InterPro; IPR003255; ATP_synt_B_c.
DR InterPro; IPR00194; ATPase_alpha_beta.
DR Pfam; PF00006; ATP_synt_ab_1.
DR Pfam; PF01038; ATP_synt_B_c_1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
KW Hydrolyase; ATP synthetase; Hydrogen ion transport.
FT CONFLICT 79 79 V > L (IN REF. 1).
FT CONFLICT 227 227 Q > R (IN REF. 4).
FT CONFLICT 500 517 DTRSGKKKSDAQEESLI -> TQAPVRRRTPAKKNL
FT (IN REF. 1).
SQ SEQUENCE 517 AA; 57749 MW; 02A7DEC571EFF7C2 CRC64;

Query Match 5.6%; Score 111.5; DB 1; Length 517;
Best Local Similarity 20.0%; Pred. No. 0.26;
Matches 82; Conservative 62; Mismatches 140; Indels 125; Gaps 18;

QY 27 LNTILSRMKERVF-----CSRTCRITTTIQDGLSVKQLGTQNVLDLIKLVENNH 76
Db 161 MNSIARGOKIPFSASGLPHNETAAQICROAGLVRP-----TKDVHD-----CHEE 206
QY 77 NFGIHEMRVSSDLPFPFASHAKYGYTFLEFAQSHLEVG-----KLANK-----YNH 121
Db 207 NFSIVFAAMGVNL-----ETARFFKQDFEENGSLERTSLFLNLANDPRIIITP 256
QY 122 RLTHMPGOYTOIASPREVVV-----DSAIRDL-AYHDEILSRMKLNQLNKDAVLII 172
Db 257 KLALTATAYLAYQTERHVLTLTDMSSYADALREVSAAREVEVPGRGYPCGYMTDLSTIV 316
QY 173 HUGTEPEGKKTLDLRFKNYQRLSDSVKARLVLENDVSVSDVODLLPLCQELNIPVLQW 232
Db 317 ERAGRVEGRNGSIQT-----PLITMPNDIHTIPDLTGYTEGIFVDROL 364
QY 233 HHINIVPGTLREGSLDMLPIPTIRETWTKGITOKOH-----YS---ESADPTA 279
Db 365 HNRGIYV-----PINVLPSLSRLMKSAIGEGMTRKDHGVDVSNOLYAKYAIKGAAMKA 418
QY 280 ISGMKRRRAHSDRYDFPFPDPTMDLMEAKEK-----EQAVFE-----LCR 320
Db 419 VVGEELSLIEDK-----LSLEFLEKFEKFTITQGGAYEDRTVFESLDQAWSLLR 466
QY 321 RY--ELQNFPCPLEIMGPEYDQTR--GYPPGAERKRLTARKRRSRKEE 365
Db 467 IYPKEMLRNTRSP-KILDEFVDRARDADADEEDPDPTRSSGKKKDSQEE 514

RESULT 3
VATB_DESSY
ID VATB_DESSY STANDARD; PRT; 463 AA.
AC 006505;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE V-TYPE ATP SYNTHASE BETA CHAIN (EC 3.6.1.34) (V-TYPE ATPASE SUBUNIT
DE H).
GN ATPB.
OS Desulfurococcus sp. (strain SY).
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC Desulfurococcus.
OX NCBI_TaxID=59822;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97320421; PubMed=9177272;
RA Shibui H., Hamamoto T., Yoshida M., Kadawa Y.;
RT "The stabilizing residues and the functional domains in the
RL hyperthermophilic V-ATPase of Desulfurococcus".
RL Biochem. Biophys. Res. Commun. 234:341-345(1997).
CC -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE. THE ARCHAEL BETA CHAIN IS A

```

CC REGULATORY SUBUNIT.
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U96487; AAB64417.1; ALT INIT.
DR InterPro: IPR003255; ATP_synth_B_c.
DR InterPro: IPR000194; ATPase_alpha_beta.
DR Pfam: PF00006; ATP_synth_ab; 1.
DR Pfam: PF01038; ATP_synth_B_c; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
DR Hydrolase; ATP synthesis; Hydrogen ion transport.
SQ SEQUENCE 463 AA; 52051 MW; 8D94195D02D5D09A CRC64;

Query Match 5.6%; Score 110; DB 1; Length 463;
Best Local Similarity 19.1%; Pred. No. 0.29;
Matches 71; Conservative 55; Mismatches 153; Indels 92; Gaps 13;

QY 27 LNTILSRMKERVPCSRCTRTTIQDGLSVKQLGTQNVLDLKLVEWNHNFQIHFMVRS 86
DB 139 MNTLVRGQLPIFGSGLPHNMLAAQIAQKAVLGEE-----EQFAVVF----- 182

QY 87 SDLFPFASHAKYGYTLE---FAQSHLEEVCKLANK-----YHRLTMHPG 128
DB 183 -----AAMGITYEANFKKSFEEGATRAVFLNLDADPAIERIIPRMALTYA 233

QY 129 QYTIQIASPREVVV-----DSAIRDL-AYHDEILSRMKLNEQLNKDAVLIHLGGTPE 179
DB 234 EYLAFDYDMQVILVILDMTNYAALREISAAREEVPGRGYPGYMTDLATIVERGRVR 293

QY 180 GKKETLDRFRKNYQRLSDSVKARLVLENDVSVSDVLDLPLCOELNIPVLWDDHHNIVP 239
DB 294 GKGSITQM-----PILTMPDDDDITHPDLTGYTEGQIVLSRELHRKGIYP 341

QY 240 GTLREGSLDMLPIPTIRETWTRKGIKQKH-----YSES---ADPTAISGMKRR 286
DB 342 -----PIDVLPSLSRMLKDGICKGRTRDHPQLAQLVAAVAGRSRLDLVAVVGEAL 395

QY 287 AHSRVDVDFPPCDPTMDLMEAKEQAVFE-ICRYEYELQNPPCPLEIMGPEVDQTRGY 345
DB 396 SETDRKYLKFADRFEREFIAQRYDEDRSIFETLDLQWELL-----AELPESELKRVKREY 450

QY 346 ---YPPGAERK 353
DB 451 ILKYHPKRR 461

RESULT 4
VATB_ARATH
ID VATB_ARATH STANDARD; PRT; 492 AA.
AC P11574; Q42205;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VACUOLAR ATP SYNTHASE SUBUNIT B (EC 3.6.1.34) (V-ATPASE B SUBUNIT)
DE (VACUOLAR PROTON PUMP B SUBUNIT) (V-ATPASE 57 KDA SUBUNIT).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89053962; PubMed=2903860;
RA Manolson M.F., Ouellette B.F.F., Fillion M., Poole R.J.;
RT "cDNA sequence and homologies of the '57-kDa' nucleotide-binding

RT subunit of the vacuolar ATPase from Arabidopsis.";
RL J. Biol. Chem. 263:17987-17994(1988).
RN [2]
RP SEQUENCE OF 241-340 FROM N.A.
RC STRAIN=CV; COLUMBIA; TISSUE=Seedling;
RA Desprez T., Anselme J., Chiapello H., Caboche M., Hofte H.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NONCATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF
CC VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.
CC -!- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,
CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04185; AAC36485.1; -.
DR EMBL; Z29126; CAA82390.1; -.
DR PIR; A31886; A31886.
DR InterPro: IPR003255; ATP_synth_B_c.
DR InterPro: IPR000194; ATPase_alpha_beta.
DR Pfam: PF00006; ATP_synth_ab; 1.
DR Pfam: PF01038; ATP_synth_B_c; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
DR Hydrolase; ATP synthesis; Hydrogen ion transport.
KW CONFLICT 303 303 R -> V (IN REF. 2).
SQ SEQUENCE 492 AA; 54739 MW; 2BF630AA79EEAC7C CRC64;

Query Match 5.5%; Score 108.5; DB 1; Length 492;
Best Local Similarity 20.0%; Pred. No. 0.41;
Matches 69; Conservative 60; Mismatches 123; Indels 93; Gaps 16;

QY 27 LNTILSRMKERV-----CSRTCRTTIQDGLSVKQLGTQNVLDLKLVEWNH 76
DB 158 MNSIARGQKIPLSAAGLPHNEIAAQICR-----QAGL--VKRL--EKTVDLLE----DH 204

QY 77 ---NFGIHFMRVSSDLFPFASHAKYGYTLFEAQSHLEEVG-----KLANK----- 118
DB 205 SEDNFAIVFAAMGVNM-----ETAQPFKRDFEENGSMERYTLFLNLANDPITRI 254

QY 119 YNHLTMHPGOVYTIASPREVVV-----DSAIRDL-AYHDEILSRMKLNEQLNKDAV 169
DB 255 ITPRIALTAEYLAECGKHVLVILTMSSYADALREVSAAEEVPGRGYPGYMTDLA 314

QY 170 LTIHLGGTFEGKKETLDRFRKNYQRLSDSVKARLVLENDVSVSDVLDLPLCOELNIPVL 229
DB 315 TIYERAGRIEGRKGSITQI-----PILTMPDDDDITHPDLTGYTEGQIVID 362

QY 230 LDWHHNVIPGTLREGSLDMLPIPTIRETWTRKGIKQKH-----YSESADPTAISG 282
DB 363 RQLHNRKQIYP-----PINVLPSSLRMLKSAIGEMTRKDSVSNQLYANYAIGKQVQA 416

QY 283 MKRRAHSRVRVDFPPCDPTMDLM-TEAKEQAVFELCRRYELQN 326
DB 417 MKAVVGEAL-----SSEDLILYLFLELDKFERKFMQCAVDTRN 454

RESULT 5
VATB_CYACA
ID VATB_CYACA STANDARD; PRT; 500 AA.
AC P48413;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE VACUOLAR ATP SYNTHASE SUBUNIT B (EC 3.6.1.34) (V-ATPASE B SUBUNIT)
DE (VACUOLAR PROTON PUMP B SUBUNIT)
OS Cyanidium caldarium.
OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC Cyanidium.
OX NCBI_TaxID=2771;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95345110; PubMed=7619836;
RA Ziegler K., Hauska G., Nelson N.;
RT "Cyanidium caldarium genes encoding subunits A and B of V-ATPase.";
RL Biochim. Biophys. Acta 1230:202-206(1995).
CC -!- FUNCTION: NONCATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF
CC VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.
CC -!- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,
CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U17101; AA85821.1; -
DR InterPro: IPR003255; ATP_synt_B_c.
DR IPR000194; ATPase_alpha_beta.
DR Pfam: PF00006; ATP_synt_ab; 1.
DR Pfam: PF01038; ATP_synt_B_c; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
KW Hydrolase; ATP synthesis; Hydrogen ion transport.
SQ SEQUENCE 500 AA; 56433 MW; 4A882382E8E9CF5 CRC64;

Query Match 5.5%; Score 108; DB 1; Length 500;
Best Local Similarity 18.7%; Pred. No. 0.46;
Matches 66; Conservative 52; Mismatches 129; Indels 106; Gaps 12;
QY 5 PREMFCDLKKIPWRGLGYACILNTILSRMKERYF-----CSRTRIT-- 48
Db 142 PEMFE-----TGSSIDVMNSIARGQKIPFLSGAGLPHEVAQAICRQCLVSTCT 193
QY 49 -IQRDGLSVKQLGTQNVLDLKLVEVNHNFPIHFMVSSDLPFPFASHAKYGVYTLFAQS 107
Db 194 LVKRSKGD-----EEDFAIVFAAMGVNM-----ETAREFRQ 224
QY 108 HLEEVG-----KLANK-----YNHRLTHPGGYTOIASPREVV-----DSA 144
Db 225 DFEENGAMERTVTLFLNLANPTIERITPRALTAFAEYLAYEKGHVLTLTDMASAYADA 284
QY 145 IRLD-AYHDEILSRMKNEQLNDAVLIHLGGTFEKKETLDRFRKNYQRLSDSVKARL 203
Db 285 LREVSAAEEVPGRGVGYMTDLATYIYERAGVRGPGSIQL-----PIL 332
QY 204 VLENDVSVSWQDILLPLCQELNIPVLVDWHHNNIPVPTLREGSLDMLPLTIPTRETWTR 262
Db 333 TPNDDITHPIPDLTGYITEQIVLSRELHAKGIYP-----PINVLPSLRMKSAIGE 386
QY 264 GITOKOHYESADPTALSGMKRRAHSRDFVDFPCCDPTMDLMTAEKKEAVF 316
Db 387 GMTKRDSDVSNQLYAAYAMGKDALAMRAV-----VGVLEALSQEDLLY 429

RESULT 6
ID VATB_ARCFU
AC 029100; STANDARD; PRT; 470 AA.
DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE V-TYPE ATP SYNTHASE BETA CHAIN (EC 3.6.1.34) (V-TYPE ATPASE SUBUNIT
DE B).
GN ATPB OR AF1167.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Hickey E.K., Peterson J.D.,
RA Ketchum K.A., Dodson R.J., Winn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Cocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE. THE ARCHAEL BETA CHAIN IS A
CC REGULATORY SUBUNIT.
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF001023; AAB90073.1; ALT_INIT.
DR TIGR: AF1167; -
DR InterPro: IPR003255; ATP_synt_B_c.
DR InterPro: IPR000194; ATPase_alpha_beta.
DR Pfam: PF00006; ATP_synt_ab; 1.
DR Pfam: PF01038; ATP_synt_B_c; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
KW Hydrolase; ATP synthesis; Hydrogen ion transport; Complete proteome.
SQ SEQUENCE 470 AA; 52874 MW; C57B2F845FB873E2 CRC64;

Query Match 5.3%; Score 104.5; DB 1; Length 470;
Best Local Similarity 20.2%; Pred. No. 0.79;
Matches 51; Conservative 45; Mismatches 93; Indels 63; Gaps 9;
QY 144 ATARDI-AYHDEILSRMKNEQLNDAVLIHLGGTFEKKETLDRFRKNYQRLSDSVKAR 202
Db 254 ALRETSAAEEVPGRGVGYMTDLATYIYERAGVRGPGSIQL-----PI 301
QY 203 VLENDVSVSWQDILLPLCQELNIPVLVDWHHNNIPVPTLREGSLDMLPLTIPTRETWTR 262
Db 302 LIMPDDITHPIPDLTGYITEQIVLSRELHAKGIYP-----PINVLPSLRMKKEGIG 355
QY 263 KITOKQH-----YESAD---PTAIGMKRRAHSRDFVDFPCCDPTMDLMTAEK 309
Db 356 EGYTREDHPQWNDOMYAAVAGVDLRLGLVAIVGEALSERDLF-----LKFA 403
QY 310 EKEQAVFELCRRYELONPPCLEI-----MCPVEDQTR-----DGYVPCAEKRLTAR 357
Db 404 DEFERRFVQOGRYEDRDIEYTLGIGWELLAMLPRELTKVERKFTKYHP-----KYR 456
QY 358 KRRSRKEEVED 369
| : : : | | |


```
QY 31 LRSKRVFCSRTCRITTTIORDGLESV-----KOL-----GTONVLDLIKVE 73
DB 1622 LKALEDRLSESAAKLAELKRAKBAQKIAIKKOLLSQMEKEEQYKKGTFESHLSELNTKL 1681
QY 74 WHNFIHGMFMR-----VSSD-----LFFPASHAKYGYTLEFAQSHLEEVGKILANKYNHRL 123
DB 1682 QEREREVHILEEKLKSVSSQSETLIVPSAKNVAAYT---EQEADSGQGVOKTYEKKI 1738
QY 124 THMPGYGTO-----IASPREVVVDASIRDLAYHDEILSRMKLNE-QLNKDAVLIHLG 175
DB 1739 SVLQRLNTEKEKLLQORGOKEKETVSSHFMRCQYQERLIKLEHAEAKOHEQDSMIGHL- 1797
QY 176 GFTEGKETLDRFKKNYQRL-----SDSVKARLVLEN--DDVSWSVODLLPLQO- 222
DB 1798 -----QEELEEKKNYSILVIAOHVEKEGKNIAQONKQNFDDVQKTLQEKELTCOI 1851
QY 223 -----ELNPLVLDDHHNINVPGLT-----EGSLDMLPLIPTRETWTTRKGTQKHVSE 273
DB 1852 LEQIKELSDCLVROKEVHRVEMBELTSKYVKLOAQOQDGRNKPTELLEENTEESKSH 1911
QY 274 SADPTAISGMKRAHSDRVDFPPCDPTWDLMTAEAKEQAVFELCR-----RYELO 325
DB 1912 LVQPKLLSNMEAQ-HNDLEFKLAGAE-----REKOKLGEIVRLQKDLRMLRKEHQ 1961
QY 326 NPPCPLEIMGPEDQTRDGYPPGAEKRLTARKRSRKEEVE 367
DB 1962 Q-----ELEILKKEYDQERE-----EKIKOEQEDLE 1987

RESULT 10
VATB_HELVI
ID VATB_HELVI STANDARD; PRT: 494 AA.
AC P31410;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VACUOLAR ATP SYNTHASE SUBUNIT B (EC 3.6.1.34) (V-ATPASE B SUBUNIT)
DE (VACUOLAR PROTON PUMP B SUBUNIT).
GN VHA55.
NCBI_TaxID=7102;
SEQUENCE FROM N.A.
TISSUE=Midgut;
MEDLINE=92027804; PubMed=1834020;
Gill S.S., Ross L.S.;
RT "Molecular cloning and characterization of the B subunit of a
RT vacuolar H(+)-ATPase from the midgut and Malpighian tubules of
RT Helicoverpa virescens.";
RL Arch. Biochem. Biophys. 291:92-99(1991).
CC -!- FUNCTION: NONCATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF
CC VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.
CC -!- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,
CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; S61797; AAB20098.1;
DR PIR; S18395; S18395.
```

```
DR InterPro: IPR003255; ATP_synt_B_c.
DR InterPro: IPR000194; ATPase_alpha_beta.
DR Pfam: PF00006; ATP_synt_ab; 1.
DR Pfam: PF01038; ATP_synt_B_c; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
KW Hydrolase; ATP synthesis; Hydrogen ion transport.
SQ SEQUENCE 494 AA; 54893 MW; 2D20C581D8AE4C76 CRC64;

Query Match 5.2%; Score 103; DB 1; Length 494;
Best Local Similarity 19.7%; Pred. No. 1.1;
Matches 61; Conservative 48; Mismatches 114; Indels 86; Gaps 13;

QY 7 EMFCLDKPI-PW-----RGLGYACILNLTLSMKERVF-----CSRTCR 45
DB 132 DFLDIQOGIPNWSKIYPEEMIQTCISAIDVNSIARQKIPISAGLPHNEIAAQICR 191
QY 46 ITTIORDGLESVKQGTQNVLDLKLVEHNNFNGIIFMRVSSDLFPFASHAKYGYTLEFA 105
DB 192 -----QAGL--VKIPG-KSVLD-----DHEDNFAIVFAAMGVNM-----ETARFF 228
QY 106 QSHLEEVCKLANK-----YNHRLTMHPQOYTOIASPREVVV-----D 142
DB 229 KODFEENGSMENVCLFLNLANDPRTIERIITPRALATAAEFLAYQCEKHVILVLTDMSSYA 288
QY 143 SAIRDL-AVHDEILSRMKLNEQLNKDAVLIITHLGTFEKGKETLDRFRKNYQRLSDSVKA 201
DB 289 EALREVSAAAREVFORGFPFGYMYTDLATIERAGRVGRCNGSITQI-----P 336
QY 202 RLVLENDVSWSVODLLPLCOELNPLVLDWHHHNINVPGLTREGSLDMLPLIPTRETWT 261
DB 337 ILTMPNDITTHIPDLTGIVTEGGIYVDRQLHNRQIYP-----PVNPLSLSLRMSKSAI 390
QY 262 RKGITOKOH 270
DB 391 GECMTRKDH 399

RESULT 11
VATB_MANSE
ID VATB_MANSE STANDARD; PRT: 494 AA.
AC P31401;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VACUOLAR ATP SYNTHASE SUBUNIT B (EC 3.6.1.34) (V-ATPASE B SUBUNIT)
DE (VACUOLAR PROTON PUMP B SUBUNIT).
GN VHA55.
NCBI_TaxID=7130;
SEQUENCE FROM N.A.
TISSUE=Midgut;
MEDLINE=92379093; PubMed=1387326;
Novak F.J., Graf R., Waring R.B., Wolfersberger M.G., Wiecezorek H.,
Harvey W.R.;
RT "Primary structure of V-ATPase subunit B from Manduca sexta midgut.";
RT Biochim. Biophys. Acta 1132:67-71(1992).
CC -!- FUNCTION: VACUOLAR ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS. THE B SUBUNIT
CC IS NONCATALYTIC BUT COMBINES WITH OTHER SUBUNITS TO FORM THE
CC CATALYTIC COMPLEX. V-ATPASE IS RESPONSIBLE FOR ENERGIZING
CC ELECTROPHORETIC K+/2H+ ANTIPORT BY GENERATING A TRANSMEMBRANE
CC VOLTAGE OF MORE THAN 200MV.
CC -!- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,
CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC
```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; X64354; CAA45706.1; -
 DR PIR; S19206; S19206.
 DR PIR; S24387; S24387.
 DR InterPro; IPR003255; ATP_synt_B_C.
 DR InterPro; IPR000194; ATPase_alpha_beta.
 DR Pfam; PF00006; ATP_synt_ab; 1.
 DR Pfam; PF01038; ATP_synt_B_C; 1.
 DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
 DR Hydrolase; ATP synthesis; Hydrogen ion transport.
 KW SEQUENCE 494 AA; 54904 MW; 8A8A3A2C04A9CFD6 CRC64;
 SQ

Query Match 5.2%; Score 103; DB 1; Length 494;
 Best Local Similarity 19.7%; Pred. No. 1.1;
 Matches 61; Conservative 48; Mismatches 114; Indels 86; Gaps 13;

QY 7 ENFDCLDKPI-PW-----RGLGYACLTITLSMKERVF-----CSRTCR 45
 DB 132 DFLDLOQPINPWSRIYFEEMLOTGISAIDVWNSIARGKIPISAGLPHNEIAAQICR 191
 QY 46 ITTIQDGLSVKQLGTONVLDLKLVEWNNHFGIHPMRVSSDLFPFASHAKYGYTLEFA 105
 DB 192 -----QAGL--VKIPG-KSVLD-----DHEDNFAIVFAAMGVNM-----ETARFF 228
 QY 106 OSHLEEVCKLANK-----YNRLTMHFGQYTOIASPREVV-----D 142
 DB 229 KQDFENGSMENVCLFLMLANDPTIERIITPRALTAETAEFLAYQCEKHVLVILTDMSSYA 288
 QY 143 SAIRDL-AYHDEILSRMKLNQKDAVLIHLGGTFEGKKTETLDRFRKNYORLSDSVKA 201
 DB 289 EALREVSAAAREVPGRCGFGMYTDLATIVERAGRVEGRGSGITQI-----P 336
 QY 202 RLVLNDDSVSVODLLPLCOELNIPVLVDWHHNIIVPGTLREGSLDMLPIPTIRTW 261
 DB 337 ILTMNDNDIITHIPDLTGTYITEGQIYVDROLHNRQIYP-----PVNVLPSLRMKSAI 390
 QY 262 RKGITQKQH 270
 DB 391 GEGMTRKDH 399

RESULT 12
 VABI_BOVIN STANDARD; PRT; 513 AA.
 AC P31407;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE VACUOLAR ATP SYNTHASE SUBUNIT B, KIDNEY ISOFORM (EC 3.6.1.34) (V-
 DE ATPASE B1 SUBUNIT) (VACUOLAR PROTON PUMP B ISOFORM 1) (ENDOMEMBRANE
 DE PROTON PUMP 58 KDA SUBUNIT).
 GN ATP6B1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 504-513.
 RC TISSUE=Kidney;
 EX MEDLINE=9222822; PubMed=1373501;
 RA Nelson R.D., Guo X.-L., Masood K., Brown D., Kalkbrenner M.,
 RA Gluck S.;
 RT "Selectively amplified expression of an isoform of the vacuolar H(+)-
 RT ATPase 56-kilodalton subunit in renal intercalated cells.";

RL Proc. Natl. Acad. Sci. U.S.A. 89:3541-3545(1992).
 CC -!- FUNCTION: NON-CATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF
 CC VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
 CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.
 CC -!- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
 CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,
 CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
 CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).
 CC -!- SUBCELLULAR LOCATION: PLASMA MEMBRANE OF THE RENAL INTERCALATED
 CC CELLS.
 CC -!- TISSUE SPECIFICITY: KIDNEY CORTEX AND MEDULLA.
 CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; M88591; AAA30394.1; -
 DR PIR; C44138; C44138.
 DR InterPro; IPR003255; ATP_synt_B_C.
 DR InterPro; IPR000194; ATPase_alpha_beta.
 DR Pfam; PF00006; ATP_synt_ab; 1.
 DR Pfam; PF01038; ATP_synt_B_C; 1.
 DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
 DR Hydrolase; ATP synthesis; Hydrogen ion transport; Multigene family.
 KW SEQUENCE 513 AA; 56746 MW; 4C6DBE4E12570A25 CRC64;
 SQ

Query Match 5.2%; Score 102; DB 1; Length 513;
 Best Local Similarity 19.7%; Pred. No. 1.4;
 Matches 59; Conservative 49; Mismatches 108; Indels 84; Gaps 12;

QY 5 PREMFDCLDKPIPWRCRCLGYACLTITLSMKERVF-----CSRTCRITTTQROGL 54
 DB 160 PEMIETGISPID-----VWNSIARGKIPISAGLPHNEIAAQICR-----QAGL 206
 QY 55 ESKVQLGTQNVLDLKLVEWNNHFGIHPMRVSSDLFPFASHAKYGYTLEFAOSHLEEVGK 114
 DB 207 --VKK--SKAVLDY-----HDDNFAIVFAAMGVNM-----ETARFFKSDFEQNGT 247
 QY 115 LANK-----YNRLTMHFGQYTOIASPREVV-----DSAIRDL-AY 150
 DB 248 MGNVCLFLMLANDPTIERIITPRALTAETAEFLAYQCEKHVLVILTDMSSYAALREVSAA 307
 QY 151 HDEILSRMKLNQKDAVLIHLGGTFEGKKTETLDRFRKNYORLSDSVKARLVLENDV 210
 DB 308 REEVPGRRCGFGMYTDLATIVERAGRVEGRGSGITQI-----PILTMNDI 355
 QY 211 SWSVODLLPLCOELNIPVLVDWHHNIIVPGTLREGSLDMLPIPTIRTKGITQKQH 270
 DB 356 THEIPDLTGITGTYITEGQIYVDROLHNRQIYP-----PINVLPSLRMKSAICEGTRKDH 409

RESULT 13
 VABI_CHICK STANDARD; PRT; 453 AA.
 AC P49712;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE VACUOLAR ATP SYNTHASE SUBUNIT B (EC 3.6.1.34) (V-ATPASE B SUBUNIT)
 DE (VACUOLAR PROTON PUMP B SUBUNIT).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-WHITE LECHORN: TISSUE-Bone;
RX MEDLINE=9536982; PubMed=7642089;
RA Bartkiewicz M., Hernando N., Reddy S.V., Roodman G.D., Baron R.;
RT "Characterization of the osteoclast vacuolar H(+)-ATPase B-subunit.";
RL Gene 160:157-164(1995).
CC -!- FUNCTION: NONCATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF
CC VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.
CC -!- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,
CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U20766; AAA82983.1; ..
DR InterPro: IPR003255; ATP_synt_B_c.
DR InterPro: IPR000194; ATPase_alpha_beta.
DR Pfam: PF00006; ATP_synt_ab; 1.
DR Pfam: PF01038; ATP_synt_B_c; 1.
DR Pfam: PF01038; ATP_synt_B_c; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
KW Hydrolyase; ATP synthetase; Hydrogen ion transport; Multigene family.
SQ SEQUENCE 453 AA; 50225 MW; 463885F93DFEC962 CRC64;

Query Match 5.18; Score 101.5; DB 1; Length 453;
Best Local Similarity 19.28; Pred. No. 1.3;
Matches 61; Conservative 55; Mismatches 111; Indels 91; Gaps 14;
QY 2 DHAP-----REMFDCDKPWPWRGL-----GYAC---LNTILSRMKERVF----- 39
DB 81 DRGPAVLAEDFLDIMGQIPNQCRIYPEEMTQTGISAIDGNSTARGOKTIPISAAAGLPH 140
QY 40 ---CSRTCRITTIQDGLSVKOLGTQNVLDLIKLVENHNFHGMVSDLPFFASHA 96
DB 141 NEIAAQICR---QAGL--VKK--SKDMDYSE-----ENFAVFAAMGVNM----- 179
QY 97 KYGYTFLEAQSHPLEVGKLANK-----YHRLTHMHPQYQTIASPREVVV 141
DB 180 ---ETAREFKSDFENGSMQNVCLFLNLANDPTIERTIIPRLALTAEFLAYQCEKHVLY 236
QY 142 -----DSAIRDL-AYHDEILSRMKLNQKNDKDAVLIHLGCTFECKKTLDRFRKNY 192
DB 237 ILTDMSSVAEALREVSAAAREEVPGRGPGYMYTDLATYIYERAGRVGRCNSITQI---- 292
QY 193 QRLSDSVKARVLENDVSVQDILLPLCQELNPLVDWHHHNIVPTGLREGSLDMLPL 252
DB 293 -----PILTMPNDITHTIPDLTGITGTEGQIVVDRLHNRQIYP-----PINVLPS 338
QY 253 IPTIRETWKRGITOKOH 270
DB 339 USRLMKSAIGCWTRKH 356

RESULT 14
ID VABT_PYRAB STANDARD; PRT; 465 AA.
AC Q90X08;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE V-TYPE ATP SYNTHASE BETA CHAIN (EC 3.6.1.34) (V-TYPE ATPASE SUBUNIT B).
DE ATP OR PAB1186.
GN Pyrococcus abyssi.
OS Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OC

OX NCBI_TaxID=29292;
RN {1}
RP SEQUENCE FROM N.A.
RC STRAIN-ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON GRADIENT ACROSS THE MEMBRANE. THE ARCHAEL BETA CHAIN IS A REGULATORY SUBUNIT.
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ248288; CAB50665.1; ..
DR InterPro: IPR003255; ATP_synt_B_c.
DR InterPro: IPR000194; ATPase_alpha_beta.
DR Pfam: PF00006; ATP_synt_ab; 1.
DR Pfam: PF01038; ATP_synt_B_c; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
KW Hydrolyase; ATP synthetase; Hydrogen ion transport; Complete proteome.
SQ SEQUENCE 465 AA; 52224 MW; 873F500BD4C5EDEF CRC64;

Query Match 5.11; Score 101.5; DB 1; Length 465;
Best Local Similarity 19.28; Pred. No. 1.3;
Matches 59; Conservative 46; Mismatches 119; Indels 83; Gaps 12;
QY 27 LNTILSRMKERVFCSRTCRITTIQDGLSVKOLGTQNVLDLIKLVENHNFHGMVRS 86
DB 141 MNTLVRCQKLPJESG-----SGLPH-NOLAAQ-IARQAKVLGESEFAVVF---- 184
QY 87 SDLPFPASHAKYGYTLE---FAQSHLEEVGKLANK-----YHRLTHMHPG 128
DB 185 -----AAGITVEANFFPKSFEETGAIERAVLFLNLADDPAIERITPRMALTV 235
QY 129 QYQTIASPREVVV-----DSAIRDL-AYHDEILSRMKLNQKNDKDAVLIHLGCTFE 179
DB 236 EYLAFDYDMQVLLVDMNYCEALREISAAREEVPGRGPGYMYTDLATYIYERAGR 295
QY 180 GKKETLDRFRKNYORLSVYKARVLENDVSVQDILLPLCQELNPLVDWHHHNIVP 239
DB 296 GKKSITQM-----PILTMPDDITHTIPDLTGITGTEGQIVLSRELHRRGIYP 343
QY 240 GTLREGSLDMLPLIPTIRETWKRGITOKOH-----YSES---ADPTAISGMKRR 286
DB 344 -----PIDVLPSSLRMKDGIGKGRTRDPOLAOOLYAAAYAEGRSLDLVAVVGEAL 397
QY 287 AHSDRVF 293
DB 398 SETDKKY 404

RESULT 15
ID VABT_CAEEL STANDARD; PRT; 491 AA.
AC Q19626;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE VACUOLAR ATP SYNTHASE SUBUNIT B (EC 3.6.1.34) (V-ATPASE B SUBUNIT) (VACUOLAR PROTON PUMP B SUBUNIT).
DE F20B6.2.
GN Caenorhabditis elegans.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae.
OC Rhabditidae; Peloderinae; Caenorhabditis.

```
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Minx P.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NONCATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF
CC VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS (BY SIMILARITY).
CC -!- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,
CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN) (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U41015; AAA82311.1; -
DR WormPep; F20B6.2; CE04424.
DR InterPro; IPR003255; ATP_synt_B_c.
DR Pfam; PF00006; ATP_synt_ab; 1.
DR Pfam; PF01038; ATP_synt_B_c; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
KW Hydrolyase; ATP synthesis; Hydrogen ion transport.
SQ SEQUENCE 491 AA; 54750 MW; EDF4D13B3D34716 CRC64;

Query Match 5.13; Score 101; DB 1; Length 491;
Best Local Similarity 18.43; Pred. No. 1.6;
Matches 75; Conservative 66; Mismatches 158; Indels 108; Gaps 16;

Qy 7 EMFCDLKP-I-PW-----RGLGYACLNPIILSRKERVPCSTCRITTIQRDGL- 54
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 55 --ESVKQLGTONVLDLKLVEWNH---NFGIHFMRVSSDLFPFASHAKYGYTLEFAQSHL 109
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 181 HNEIAAQIVROG--GLVQLPDRPHEQTNEAIVFAAMGVNM-----ETARFFKQDF 228
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 110 EEVGKLANK-----YNHRLTMHPGQYQIASPREVVY-----DSAIR 146
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 229 EENGSMENVCLFLNANDPTTERTIITPRIALTSAEFLAYOCKKHVIVLVLTDMSSYAEALR 288
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 147 DL-AYHDEILSRMKLNQKNKDAVLIILHGGTFEGKKEITLDRFRKNYQRLSDSVKARLVL 205
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 289 EVSAAAREEVPGRGFGPGVMYTDLATIVERAGRVGRGDSITQI-----PILTM 336
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 206 ENDDVSVSQDLLPLCOELNIPVLVDHWHHNIIVPGTLREGSLDMLPIPTIRETWTRKGI 265
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 337 PNDDITHPIPLDTGYITEGQIYVDVDRQLHNLIIYP-----PINVLPSLSRLMKSATGEGM 390
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 266 TOKOH-----YSESAD---PTAISGMKRAHSDRVDFPPCDPTMDLMEAKEKE 312
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 391 TREHSDVSNOLYACYAIGKDVQAMKAVGVEEALSSDDLLILEFLTKFEKNFTIQGHYEN 450
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 313 QAVFE-----LCRRYELQNPCCPLEIMGYDQTRDGYPPGAEK 352
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 451 RSVFESLDIGWQLLRIF-----PREMLKRIPESTLEKYYPRGGAK 490
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Search completed: January 15, 2002, 14:07:07
Job time: 936 sec
```



```

; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 793
; TYPE: prt
; ORGANISM: Thermotoga maritima
US-08-676-444-5

```

Query Match 4.5%; Score 90; DB 4; Length 793;
Best Local Similarity 22.9%; Pred. No. 1.1;
Matches 94; Conservative 64; Mismatches 138; Indels 114; Gaps 25;

Qy	39	FCSTRCITTIQRDGLSVKQOLCTQNVLDLIKLVENHNFI-----HFNRVSSDLEP	91
Db	130	YCTVFCDSGEVLVHESDSE--QETDLQK-----NYSQITCPEHLKSSLKREPF	180
Qy	92	FASHAKYGYTL-BFAOSHLEEVCKLAN-KYNHRLTMHP-----GOVTOIA---	134
Db	181	---GVVTETISEWYFSDLESEKAYNLKDIHFELSPILAKALAALIYVYKVTMIGEDL	236
Qy	135	-----SPRE-VVVDSEA-IRDLAYHDEILSRMKLEQNLKDAVLIIHLGDTFEGCKET	184
Db	237	NLKPLPLISQDYMILDSATVENSMTPGDRGK-NLFDVLNVTETPM-----GARLGKKWI	291
Qy	185	L-----DRFRKNYQRLSDSVKARVLVLENDVSW--SVODLLPLQOELNIPVLVDMWHHNIVP	239
Db	292	LHPLVDR-----KQIERLKAVERLVNDRVSLSEEMNLISNVRDVE-RIVSRVEYNRSVP	345
Qy	240	G---TLREGSLDMLPIPTTRETWTRKGITOKOHYSESA-----DPTAISG----	282
Db	346	RDLVALRE---TLEIIPKLINEVLSTFGVEPKLAFPEGLVLDLLRKAIEDDPVSGPEGKV	401
Qy	283	MKRRHSDRVDFPPCPOPTMDLM-----TEAKEQAVFELCR-----RYELQCN	326
Db	402	IKRGFSSE-----LDEYRDLLHAERLKEFEERERTGTQKLRVGVNQVFGYIEV	454
Qy	327	PPCPLIMGPEYDQTR-----DGYPPGAEKRLTARKRRSRKEEVEDEK	371
Db	455	TKANLDKIPDDYERKOTLVNRSERFITP--ELKKEFTKIMAAKERIELEK	502

RESULT 5
US-08-952-127-12

```

US-08-952-127-12
: Sequence 12, Application US/08952127
: Patent No. 6211336
: GENERAL INFORMATION:
:   APPLICANT: Shiloh, Yosef
:   APPLICANT: Tagle, Danilo A.
:   APPLICANT: Collins, Francis S.
:   TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE
:   NUMBER OF SEQUENCES: 24
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: Kohn & Associates
:     STREET: 30500 No. 6211336thwestern Hwy., Suite 410
:     CITY: Farmington Hills
:     STATE: Michigan
:     COUNTRY: U.S.
:     ZIP: 48334
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: PatentIn release #1.0, Version #1.30
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/952,127
:     FILING DATE:
:     CLASSIFICATION: 530
:   ATTORNEY/AGENT INFORMATION:
:     NAME: Kohn, Kenneth I.
:     REGISTRATION NUMBER: 30,995
:     REFERENCE/DOCKET NUMBER: 2290.00029
:   TELECOMMUNICATION INFORMATION:

```

```

: TELEPHONE: 810-539-5050
:
: TELEFAX: 810-539-5055
:
: INFORMATION FOR SEQ ID NO: 12:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 3066 amino acids
:
: TYPE: amino acid
:
: STRANDEDNESS: single
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: ORIGINAL SOURCE:
:
: ORGANISM: Mus musculus
:
: US-08-952-127-12

```

[illegible]

RESULT

US-08-966-389-4
; Sequence 4, Application US/08966389
; Patent No. 5843748
GENERAL INFORMATION:
; APPLICANT: Tetsuya NAKADA
; APPLICANT: Michio KUROTA
; APPLICANT: Hiroto CHAEN
; APPLICANT: TOSHIO MIYAKE
; TITLE OF INVENTION: TREHALOSE PHOSPHO
; TITLE OF INVENTION: AND USES
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEWMARK
STREET: 419 Seventh Street, N.W.,
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0,
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/966,389
; FILING DATE: 07-NOV-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 311,232/1999
; FILING DATE: 8-NOV-1996
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 61.716/1997
; FILING DATE: 3-MAR-1997

```
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: NAKADA-6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 773 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-966-389-4

Query Match          4.5%; Score 88.5; DB 2; Length 773;
Best Local Similarity 20.2%; Pred. No. 1.5;
Matches 50; Conservative 41; Mismatches 85; Indels 71; Gaps 10;

QY 98 YGYTLEFAQSHL-EEVGKL-----ANKYNHRLTMHPGOYTQIASPREVVVDSAIRDLAYHD 152
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 461 YGCEILFETAREWEDLGAVIPLKGNKFCINCVTGPDEYALVDNNAYTNMAKNMLEYAY 520

QY 153 EILSRMKLNEQLNKDAVLIILHGGTFEGKKEITLDRFRKNYORLSDSVKARLVLEDD-VS 211
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 521 DTANKM-----KKEVPQKYOK-----VASKLNKLDDEIVA 550

QY 212 W---SVQDLLPLCOELNI-----PLVLDWHHHNIVPCTLRG 245
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 551 WKKAADNMVLPYSKELDIIPQDSDFLYKERITVDEIPEDQFPLLLHWHYLNIRYQICK- 609

QY 246 SLDLMLPIPTIRETWRTKGITOKHYSE-----SADPTAISGMKRR-AHSDRVFDEPP 297
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 610 QPDVLLMFLOREKFTKDELKNYDYEPITTHDSSLSPAIFSIILANEIGYTDKAYKYM 669

QY 298 CDPITMDL 304
   |||
Db 670 MTARMDL 676

RESULT 7
US-09-103-509-4
; Sequence 4, Application US/09103509
; Patent No. 5876975
; GENERAL INFORMATION:
; APPLICANT: Tetsuya NAKADA
; APPLICANT: Michio KUBOTA
; APPLICANT: Hiroto CHAEN
; APPLICANT: Toshio MIYAKE
; TITLE OF INVENTION: TREHALOSE PHOSPHORYLASE, ITS PREPARATION
; TITLE OF INVENTION: AND USES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,509
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/966,389
; FILING DATE: 07-NOV-1997
; APPLICATION NUMBER: JP 311,232/1996
```

```
; FILING DATE: 8-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 61,716/1997
; FILING DATE: 3-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: NAKADA-6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 773 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-103-509-4

Query Match          4.5%; Score 88.5; DB 2; Length 773;
Best Local Similarity 20.2%; Pred. No. 1.5;
Matches 50; Conservative 41; Mismatches 85; Indels 71; Gaps 10;

QY 98 YGYTLEFAQSHL-EEVGKL-----ANKYNHRLTMHPGOYTQIASPREVVVDSAIRDLAYHD 152
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 461 YGCEILFETAREWEDLGAVIPLKGNKFCINCVTGPDEYALVDNNAYTNMAKNMLEYAY 520

QY 153 EILSRMKLNEQLNKDAVLIILHGGTFEGKKEITLDRFRKNYORLSDSVKARLVLEDD-VS 211
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 521 DTANKM-----KKEVPQKYOK-----VASKLNKLDDEIVA 550

QY 212 W---SVQDLLPLCOELNI-----PLVLDWHHHNIVPCTLRG 245
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 551 WKKAADNMVLPYSKELDIIPQDSDFLYKERITVDEIPEDQFPLLLHWHYLNIRYQICK- 609

QY 246 SLDLMLPIPTIRETWRTKGITOKHYSE-----SADPTAISGMKRR-AHSDRVFDEPP 297
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 610 QPDVLLMFLOREKFTKDELKNYDYEPITTHDSSLSPAIFSIILANEIGYTDKAYKYM 669

QY 298 CDPITMDL 304
   |||
Db 670 MTARMDL 676

RESULT 8
US-09-102-644-4
; Sequence 4, Application US/09102644
; Patent No. 5910436
; GENERAL INFORMATION:
; APPLICANT: Tetsuya NAKADA
; APPLICANT: Michio KUBOTA
; APPLICANT: Hiroto CHAEN
; APPLICANT: Toshio MIYAKE
; TITLE OF INVENTION: TREHALOSE PHOSPHORYLASE, ITS PREPARATION
; TITLE OF INVENTION: AND USES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/102,644
; FILING DATE:
; CLASSIFICATION:
```


;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/235,836C
; FILING DATE: 29-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,191
; FILING DATE: 01-11-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Bogosian, Margaret C.
; REGISTRATION NUMBER: 25,324
; REFERENCE/DOCKET NUMBER: BNL93-28A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 282-7338
; TELEFAX: (516) 282-3729
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 693 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-235-836C-72

Query Match 4.4%; Score 88; DB 4; Length 693;
Best Local Similarity 17.9%; Pred. No. 1.4;
Matches 62; Conservative 72; Mismatches 133; Indels 80; Gaps 18;
QY 71 LVEWHN--GFIHFM-----RVSSDLFPFASHAKY-----GY---TLEF 104
Db 67 LFNNSNSYKGYFNFIDDDQKKASVDFTSSKSDLSINLRILTYLIKSFY 126
QY 105 AQSHLEEVGKLANKYHRLTMHPQYTT-QIASPREVVVDSAIRLAYHDEILSRMKNEQ 163
Db 127 ERSSAELIAK-----ITTHNAVYRGDLNYKFEFIESALKSLTKENAGLSRV-YSQW 178
QY 164 LNKDAVLIHLGGTFEGKKT---LDRFKNYQLRSLSVKARLVLEND-----DVSWS 213
Db 179 AKQTQIFPLKKNILSGKIESDIDISL-----VTDKVAALLSENAEAGVNFARDIT-D 231
QY 214 VQDLLPLCQELNIPVLVDWHHNIVPGTLREGSLDLPLTIPTIETWTKGITOQOHSE 273
Db 232 IGETHKADQDKIDELD-----NVHSDSNITETIENLDQLEK--ATDEERKE 280
QY 274 SADPTAIGMKRRASRDVFPDPCDPTMDLMIE-----AKEK-EQAVFELCRRYEL 324
Db 281 IESQVDAKKKQKEELDKKAIIDLAKQOKLDFSEDNLDIQRTVREKIQEDINEINKE--- 337
QY 325 ONPPCPLEIMGPEYD---OTRDGYYPGCAEKKRLTARKRRSRKEVEE 368
Db 338 KNLKPGDVSSPKVDKQLQIKESL--EDLQELKETSDEKOKREIK 382

RESULT 11
US-08-147-812-5
; Sequence 5, Application US/08147812
; Patent No. 5766909
; GENERAL INFORMATION:
; APPLICANT: Xie, Qiao-wen
; APPLICANT: Nathan, Carl F.
; APPLICANT: Mumford, Richard A.
; APPLICANT: Calaycay, Jimmy Ramos
; TITLE OF INVENTION: DNA Encoding Inducible Nitric Oxide Synthase
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 East Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; OPERATING SYSTEM: Macintosh Centris650

;
; OPERATING SYSTEM: Macintosh 7.0.1
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,812
; FILING DATE: NO. 5766909 Available
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/841,641
; FILING DATE: 02-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wallen, John W III
; REGISTRATION NUMBER: 35,403
; REFERENCE/DOCKET NUMBER: 186581A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3905
; TELEFAX: (908) 594-4720
; TELEX: 138825
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1144 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-147-812-5

Query Match 4.4%; Score 88; DB 1; Length 1144;
Best Local Similarity 17.3%; Pred. No. 3.3;
Matches 46; Conservative 38; Mismatches 90; Indels 92; Gaps 9;
QY 37 RVCSRTCRITTTQRDGL-----ESVQLQGTQNVLDLIKLVENHNFG 79
Db 562 KVMCDQYKASTLEEQLLLVVTSTFGNGDCPSNGQTLKK-----SLFMLRELNHTFR 614
QY 80 IHPMYSSDLFPFASHAKYCYTLEFAQSHL-----EEVG----- 113
Db 615 YAVEGLGSSWIP--QCAFADHIDQKLSHLGASQIAPTGEDELSDQEDAFRSNAVQTFR 672
QY 114 -----KLANKNYHRLTMHPQYTTQIASPREVVVDSAIRLAYHDEILSRMK 159
Db 673 AACETEDVRSKHILQIPKRFSTNATWEPQYRLIQSPFLDLNRAISSIHAKNVFTMLRK 732
QY 160 LNEQLNKD-----AVLIHLGGTFEGK-----ETLDRFKNYQLRSLSVKARL----- 203
Db 733 SOONLOSEKSSRTLLVLQV--TFEGSRGPSYLPGEHLGIFPGNQTLVQGILERVVDCPT 790
QY 204 -----VLENDVSVSVQDILLPLC 221
Db 791 PHOTVCLEVLDESGSYWVKDKRLPPC 816

RESULT 12
US-08-319-866-12
; Sequence 12, Application US/08319866
; Patent No. 5929223
; GENERAL INFORMATION:
; APPLICANT: Tully, Timothy P.
; APPLICANT: Yin, Jerry C.
; APPLICANT: Regulski, Michael
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF GENES
; TITLE OF INVENTION: ASSOCIATED WITH LONG-TERM MEMORY
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/319,866
FILING DATE: 7-OCT-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: CSHL94-03
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1144 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-319-866-12

Query Match 4.4%; Score 88; DB 2; Length 1144;
Best Local Similarity 17.3%; Pred. No. 3.3;
Matches 46; Conservative 38; Mismatches 90; Indels 92; Gaps 9;

QY 37 RVFCSTCRITTIQDGL-----ESVKQLGTQNVLDLIKLVENHNFG 79
Db 562 KVVCMQDYKASTLEEOLLLVVTFTFGNGDCPSNGQTLKK-----SLFMLRELNHTFR 614
QY 80 IHFMRVSSDLFPFASHAKYGYTLEFAQSHL-----EEVG----- 113
Db 615 YAVFGLGSSMYP--QCAFAHDIDQKLSHGASQLAPTGEDELSCQEDAFRSWAVQTFR 672
QY 114 -----KLANKYNHRLTMHPGQYTOIASPREVVVDSAIRDLAYHDEILSRMK 159
Db 673 AACETFDVRSKHHIOPKRFTSNATWEPQOYRLIOSPEPLDNLRALSSIHAKNVFTMLRK 732
QY 160 LNEQLNKD-----AVLIHLGGTFEGKK-----ETLDRFRKNYORLSDSVKARL----- 203
Db 733 SQONLOSEKSSRTLLVQL--TFEGSRGSPYLPGEHLGIFPGNOTALVOGILERVWDCTP 790
QY 204 -----VLENDVSWSQDILLPLC 221
Db 791 PHQTVCLEVLDSEGSYVWKDKRLPPC 816

RESULT 13

US-09-123-708-2
Sequence 2, Application US/09123708
Patent No. 6146887
GENERAL INFORMATION:
APPLICANT: SCHRADER, Juergen
APPLICANT: GODECKE, Axel
TITLE OF INVENTION: DNA EXPRESSION VECTORS FOR USE IN GENE THERAPEUTIC
TITLE OF INVENTION: TREATMENT OF VASCULAR DISORDERS
FILE REFERENCE: 511169-2003
CURRENT APPLICATION NUMBER: US/09/123,708
CURRENT FILING DATE: 1998-07-28
EARLIER APPLICATION NUMBER: 08/553,503
EARLIER FILING DATE: 1996-03-01
EARLIER APPLICATION NUMBER: P4411402.8
EARLIER FILING DATE: 1994-03-31
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 1144
TYPE: PRT
ORGANISM: Cytomegalovirus
US-09-123-708-2

Query Match 4.4%; Score 88; DB 4; Length 1144;
Best Local Similarity 17.3%; Pred. No. 3.3;
Matches 46; Conservative 38; Mismatches 90; Indels 92; Gaps 9;

QY 37 RVFCSTCRITTIQDGL-----ESVKQLGTQNVLDLIKLVENHNFG 79
Db 562 KVVCMQDYKASTLEEOLLLVVTFTFGNGDCPSNGQTLKK-----SLFMLRELNHTFR 614
QY 80 IHFMRVSSDLFPFASHAKYGYTLEFAQSHL-----EEVG----- 113
Db 615 YAVFGLGSSMYP--QCAFAHDIDQKLSHGASQLAPTGEDELSCQEDAFRSWAVQTFR 672
QY 114 -----KLANKYNHRLTMHPGQYTOIASPREVVVDSAIRDLAYHDEILSRMK 159
Db 673 AACETFDVRSKHHIOPKRFTSNATWEPQOYRLIOSPEPLDNLRALSSIHAKNVFTMLRK 732
QY 160 LNEQLNKD-----AVLIHLGGTFEGKK-----ETLDRFRKNYORLSDSVKARL----- 203
Db 733 SQONLOSEKSSRTLLVQL--TFEGSRGSPYLPGEHLGIFPGNOTALVOGILERVWDCTP 790
QY 204 -----VLENDVSWSQDILLPLC 221
Db 791 PHQTVCLEVLDSEGSYVWKDKRLPPC 816

RESULT 14

US-09-123-624-2
Sequence 2, Application US/09123624
Patent No. 6149936
GENERAL INFORMATION:
APPLICANT: SCHRADER, Juergen
APPLICANT: GODECKE, Axel
TITLE OF INVENTION: DNA EXPRESSION VECTORS FOR USE IN THE GENE THERAPEUTIC
TITLE OF INVENTION: TREATMENT OF VASCULAR DISORDERS
FILE REFERENCE: 511169-2004
CURRENT APPLICATION NUMBER: US/09/123,624
CURRENT FILING DATE: 1998-07-28
PRIOR APPLICATION NUMBER: 08/553,503
PRIOR FILING DATE: 1996-03-01
PRIOR APPLICATION NUMBER: 4411402.8
PRIOR FILING DATE: 1994-03-31
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 1144
TYPE: PRT
ORGANISM: Mus musculus
US-09-123-624-2

Query Match 4.4%; Score 88; DB 4; Length 1144;
Best Local Similarity 17.3%; Pred. No. 3.3;
Matches 46; Conservative 38; Mismatches 90; Indels 92; Gaps 9;

QY 37 RVFCSTCRITTIQDGL-----ESVKQLGTQNVLDLIKLVENHNFG 79
Db 562 KVVCMQDYKASTLEEOLLLVVTFTFGNGDCPSNGQTLKK-----SLFMLRELNHTFR 614
QY 80 IHFMRVSSDLFPFASHAKYGYTLEFAQSHL-----EEVG----- 113
Db 615 YAVFGLGSSMYP--QCAFAHDIDQKLSHGASQLAPTGEDELSCQEDAFRSWAVQTFR 672
QY 114 -----KLANKYNHRLTMHPGQYTOIASPREVVVDSAIRDLAYHDEILSRMK 159
Db 673 AACETFDVRSKHHIOPKRFTSNATWEPQOYRLIOSPEPLDNLRALSSIHAKNVFTMLRK 732
QY 160 LNEQLNKD-----AVLIHLGGTFEGKK-----ETLDRFRKNYORLSDSVKARL----- 203
Db 733 SQONLOSEKSSRTLLVQL--TFEGSRGSPYLPGEHLGIFPGNOTALVOGILERVWDCTP 790
QY 204 -----VLENDVSWSQDILLPLC 221
Db 791 PHQTVCLEVLDSEGSYVWKDKRLPPC 816

Db 791 PHQTVCLVLEDSGSYVWKRLPPC 816

RESULT 15

US-08-560-005-2
; Sequence 2, Application US/08560005
; Patent No. 6001354
; GENERAL INFORMATION:
; APPLICANT: Pot, David A.
; APPLICANT: Williams, Lewis T.
; APPLICANT: Jefferson, Anne Bennett
; APPLICANT: Majerus, Philip W.
; TITLE OF INVENTION: No. 6001354el Grb2 Associating Protein and Nucleic
; TITLE OF INVENTION: Acids Encoding Therefor
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/560.005
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-0624000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 976 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-560-005-2

Query Match 4.4%; Score 87; DB 3; Length 976;

Best Local Similarity 20.4%; Pred. No. 3.3;
Matches 79; Conservative 53; Mismatches 130; Indels 126; Gaps 18;

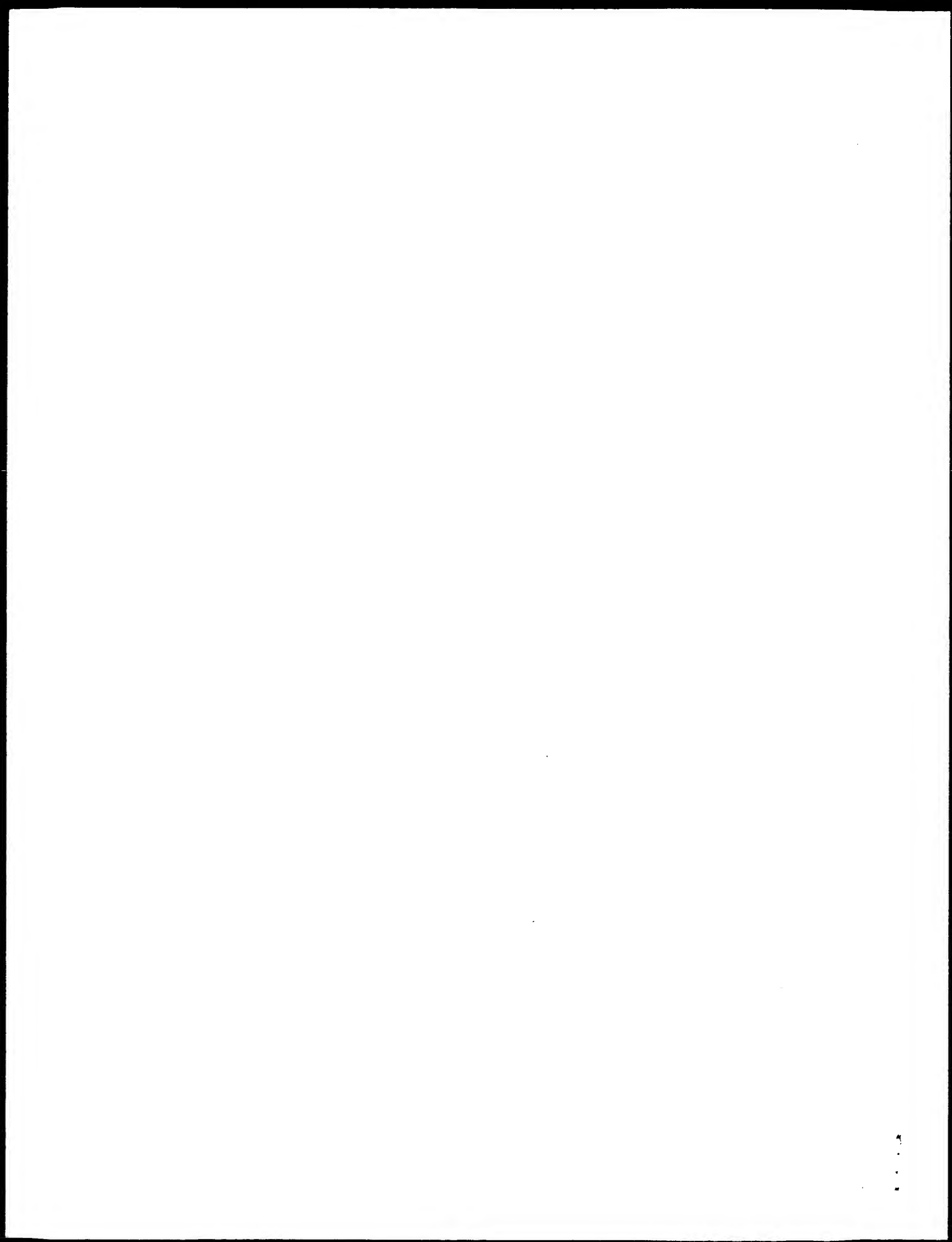
QY 24 YACLNTILRSKMERVFCSTCRIT--IQDGLSVKQLGTQNVLDLIKLVENHNFQIH 81
Db 477 YGSTDIMTSDHSPVFATFEAGVTSQFVSKNGPGTVDSQGO-----IE 519
QY 82 FMRVSSDLPPFASHAKYGYTLEFAOSHLEEV-----GKLANKYNHRLTMHPG 128
Db 520 FLRCYATL---RTKSQTKYLEFHSSCLSEFVKSOGEENEGEGELVYKFGETLP----- 572
QY 129 QYTQTASPREVVVD-----SAIRDLAYHDEILS-RMKLNF-OLNKDAVLIIH--LGG 176
Db 573 KLPILSDPEYLLDQHILISKSDSDSEYGGCGALREATELTOLPIYPLTHHGELTG 632
QY 177 TPEGKKTLDREFRKNYQRLSDSVKARLVLENDVVS-----WSVODLLPL 220
Db 633 HFQGEIKLQTSQGTREKLYDFVKT---ERDESSGPKTLKSLTSHDPMQWEVTSRAPP 688
QY 221 COELNTPVLVDWHHHNIVPGTLREGSLDMLPLIPIRETWTWKGITQKHYESADPTAI 280
Db 689 CGSSITEINFINMGVGP-----FGPPMP-----LHVKOTLSPOQQTAW 729
QY 281 SGMKRRASHDRVDFPPCDPTMDLMEAKEKEQAVFELCRRYELQNPCCP--LEIMGP 338

Db 730 S-----YDQPPKDSPLG-----PCRG-----ESPPTPGQPPLSPK- 760

QY 339 DQTRDGYYPGAEKRLTARKRRSRKEEV 366

Db 761 -----KFLPSTANRGLPPRTQESRPSDL 783

Search completed: January 15, 2002, 13:51:04
Job time: 333 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:07:26 ; Search time 45.07 Seconds
(without alignments)
239.171 Million cell updates/sec

Title: US-09-724-296-39

Perfect score: 1527

Sequence: 1 QLGLVCLTGVPEVFRFTVL.....VEAKGKEEIAALRLMAPFK 294

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	303	19.8	320	1 YWJD_BACSU	P45864 bacillus su
2	90.5	5.9	457	1 LEFA_NPVOP	O10340 orgyia pseu
3	90	5.9	488	1 DHAL_PSESP	P33008 pseudomonas
4	90	5.9	607	1 NODM_RHLIV	P08633 rhizobium l
5	89	5.8	139	1 PUAC_STRLP	P13249 streptomyce
6	89	5.8	2472	1 NCR2_MOUSE	Q9wu42 mus musculu
7	88	5.8	537	1 P4H2_MOUSE	O60716 mus musculu
8	86	5.6	1110	1 PKNK_MYCTU	P95078 mycobacteri
9	85.5	5.6	653	1 APPI_MOUSE	Q03157 mus musculu
10	84.5	5.5	304	1 HEY1_CANFA	Q9tsz2 canis famil
11	84.5	5.5	616	1 RFX5_HUMAN	P48382 homo sapien
12	84	5.5	378	1 RPSD_CLOAB	P33656 clostridium
13	82.5	5.4	271	1 TRPA_MYCIT	O68906 mycobacteri
14	82.5	5.4	650	1 APPI_HUMAN	P51693 homo sapien
15	82.5	5.4	1160	1 DP3A_ECOLI	P10443 escherichia
16	82.5	5.4	3567	1 ERY2_SACER	O03132 saccharopol
17	82	5.4	435	1 COBB_RHOCA	O68108 rhodobacter
18	81.5	5.3	967	1 HEPA_ECOLI	P23852 escherichia
19	81	5.3	506	1 GAG_SIVM1	P05894 simian immu
20	81	5.3	607	1 DNAK_STRPN	P95829 streptococ
21	81	5.3	974	1 ATXB_LEIDO	P12522 leishmania
22	81	5.3	1172	1 SYK2_MYCTU	P94974 mycobacteri
23	80.5	5.3	476	1 PCCB_RHOER	O06101 rhodococcus
24	80.5	5.3	859	1 OBP_HSVBC	P52377 bovine herp
25	80	5.2	307	1 BLAC_MYCTU	Q10670 mycobacteri
26	80	5.2	404	1 TRPB_BACST	P19868 bacillus st
27	80	5.2	612	1 UN37_CAEEL	O02482 caenorhabdi
28	80	5.2	1002	1 TNPA_ECOLI	Q00937 escherichia
29	80	5.2	1257	1 KPBA_CAEEL	P34335 caenorhabdi
30	79.5	5.2	159	1 ISPE_MYCTU	P96863 mycobacteri
31	79.5	5.2	406	1 TRPB_CAUCR	P12290 caulobacter
32	79.5	5.2	474	1 VNFK_AZOCH	P15334 azobacter
33	79.5	5.2	492	1 NYLA_FLASK	P13397 flavobacter

ALIGNMENTS

RESULT }

YWJD_BACSU

ID YWJD_BACSU STANDARD; PRT: 320 AA.

AC P45864:

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE HYPOTHETICAL 36.9 KDA PROTEIN IN ACDA 5'REGION.

GN YWJD.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RA Glaser P., de la Fuente V., Danchin A.;

RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; 249782; CA889865.1; -

DR EMBL; 299123; CAB15748.1; -

DR Subtilist; BG11309; ywjd.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 320 AA; 36896 MW; 2FD35A710648E002 CRC64;

Query Match 19.8%; Score 303; DB 1; Length 320;

Best Local Similarity 30.2%; Pred. No. 4.9e-18;

Matches 90; Conservative 50; Mismatches 116; Indels 42; Gaps 11;

Qy 16 RIVTLRYRALSPAEREAKLLDLYSSNIKTGAADYCAAHDIRLYRSLSSLPML---D 72

Db 21 KTLTFARYSKLSKTERKEALLTVTKANLRNTMRTLHYITGHGIPLYRFSSTIVPLATHPD 80

Qy 73 LAGDDTGAAVLTHLAPOLLEAGHAFTDAGVRLMLHPEQFVLNSDRPEVRESSVRMSAH 132

Db 81 VMWD-----FVTFQKEFEIGELVKTHQLTSFHPNQFTLTSPKESVTKNAVTDMAH 135

Qy 133 ARYMDGLGLA-RTPWNLL--LHGKGGRGAEALALIDLPDPVRLRLGLENDERAYS 189

Db 136 YRMLEAMGIADRSVINIHIGGAVGNKDTATAQHONIKQLPQEIKERMTLENDKTYTTE 195

Qy 190 ELLPICEATGTLPLVDAHHVVDKLPDQEDSVREWVLARATW-----QPPEQV 244

Db 196 ETLVQCEQEDVPVFDFHHFYANP--DDHADLNV--LPRMKTWERTIGLOPK----VHL 247

Qy 245 SNGIEGQDPRRHSHLADIADPSAYADVPOTE-----VEAKGKEEIAALRLM 290

```

Db 248 SS-----PKSEQAIRSHADYVDANELLPLLERFRQWNTIDFMEIAKQKDKAL--LRLM 299

RESULT 2
LEF4_NPVOP
ID LEF4_NPVOP STANDARD; PRT: 457 AA.
AC O10340;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE LATE EXPRESSION FACTOR 4.
LEF-4.
OS Orygia pseudotsugata multicapsid polyhedrosis virus (OpMPNV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohrmann G.F.;
RT "The sequence of the Orygia pseudotsugata multinucleocapsid nuclear
RT polyhedrosis virus genome.";
RL Virology 229:381-398(1997).
CC -!- FUNCTION: REQUIRED FOR LATE AND VERY LATE GENE EXPRESSION (BY
CC SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U75930; AAC59090.1; -
KW Transcription regulation.
SQ SEQUENCE 457 AA; 51151 MW; E21E06E50BAC8390 CRC64;

Query Match 5.9%; Score 90.5; DB 1; Length 457;
Best Local Similarity 23.1%; Pred. No. 3.2;
Matches 59; Conservative 34; Mismatches 91; Indels 71; Gaps 13;

Qy 24 RALSPAERAKLLDLYSSNIKTLRGAADYCAAHDIRL-----YFLSSSLFPLMLDLA 75
Db 33 RCAPAERYTLDYD--ANNVTRTADSAVSVHTNLRDERFVHMLRSSNALVPLVVRRE 90
Qy 76 DDTGA---AVLTHLAPOLLEA-----GHAFDAGVRLLMHPEOFIVLNSDRPE 120
Db 91 RETAVPHDRVSRHIA-SLIETTVYKLDGVVKFHVYMQSG-----PADRYE 136
Qy 121 VRESSVRAMSAHARVMDGLGARTPWNLLLLHGKGGKRGAEALALPDLPPVRLRLGLE 180
Db 137 --STAHHKIAALKALLGVDCARPSQNLQ-----GSDAVLA-----RVRLLE 178
Qy 181 NDERAYSPAELLPTCEATGTPLVFDAAHHVY-----HDKLPDQEDPVSREWLVRARATW-- 234
Db 179 FEGAAPAAASLDAPCELVVQMETLADHHNIAPCLPYTLLDSATP--RRTRQRIAYGA 236
Qy 235 QPPEQVQVHLSNGIE 249
Db 237 QAPD-----STGVK 245

RESULT 3
DHAL_PSESP
ID DHAL_PSESP STANDARD; PRT: 488 AA.
AC P33008;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE PROBABLE ALDEHYDE DEHYDROGENASE (EC 1.2.1.3).

```

```

GN TERPE.
OS Pseudomonas sp.
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92332528; PubMed=1629218;
RA Peterson J.A., Lu J.-Y., Geisselsoder J., Graham-Lorence S.,
RA Carmona C., Witney F., Lorence M.C.;
RT "Cytochrome P-450terp. Isolation and purification of the protein and
RT cloning and sequencing of its operon.";
RL J. Biol. Chem. 267:14193-14203(1992).
CC -!- FUNCTION: INVOLVED IN A ALPHA-TERPENEOL OXIDATION SYSTEM.
CC -!- CATALYTIC ACTIVITY: ALDEHYDE + NAD(+) + H(2)O -> ACID + NADH.
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M91440; AAA25995.1; -
KW Aldehyde dehydrogenase.
SQ SEQUENCE 483 AA; 52156 MW; 2144AE1B173C854 CRC64;

Query Match 5.9%; Score 90; DB 1; Length 488;
Best Local Similarity 23.6%; Pred. No. 3.8;
Matches 53; Conservative 32; Mismatches 82; Indels 58; Gaps 10;

Qy 54 AAHDRLRYLSSSLFPLMLDLACDDTCAAVLTHLAPOLLEA--GHAFDAGV----- 102
Db 252 AAADLK--RLT-----LELGGDAAVWADSLDAVVERVSWAFNAGQICMTAKRLY 303
Qy 103 -----RILMHPEQFIVLNSDRPEVRESSVRAMSAHARVMDGLGARTPWNLLLL 151
Db 304 IHESIYAFRDKLVAYAQNVVICDGSQPGVTMGPLQNAKQFAKVNSLIDAVRQGRGLIE 363
Qy 152 HGG-KGGRGAELALIPDLDPVRLRLGLENDERAYSPELLPICEATGTPLVFDAAHHV 210
Db 364 CGOMRGDGYFLPTFTVLPD-----ESAPEVVEAFGP--LLPLL----- 402
Qy 211 VHDKLPDQEDPVSREWLVRARATWQPPQVHLS--NGIEGQDR 254
Db 403 ---KPRDVE--VIERVNAARTGWLAAAGSVGLRSAGHGLQGPSTR 442

RESULT 4
NODM_RHLV
ID NODM_RHLV STANDARD; PRT: 607 AA.
AC P08633;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE [ISOMERIZING]
DE (EC 2.6.1.16) (G6P) (MODULATION PROTEIN M).
GN NODM.
OS Rhizobium leguminosarum (biovar viciae).
OC Plasmid sym. pRL11.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

```


[1]
 RP SEQUENCE FROM N.A. (ISOFORM ALPHA AND ISOFORM BETA).
 RC TISSUE-Spleen, and Brain;
 RX MEDLINE-9917894; PubMed-10077563;
 RA "Unique forms of human and mouse nuclear receptor corepressor SMRT.";
 RT Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).
 RL [2]
 RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
 RC TISSUE-Embryo;
 RX MEDLINE-9919215; PubMed-10097068;
 RA Park E.J., Schreier D.J., Yang M., Li H., Li L., Chen J.D.;
 RT "SMRT, a silencing mediator for retinoid and thyroid hormone
 receptor-extended isoform that is more related to the nuclear
 receptor corepressor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).
 CC -!- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
 CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
 CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
 CC -!- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
 CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
 CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE
 CC ABSENCE OF LIGAND, AND MAY STABILIZE THEIR INTERACTION WITH TFIIIB.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: UBIQUITOUS. ALSO WIDELY EXPRESSED IN EARLY
 CC EMBRYOS.
 CC -!- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
 CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2
 CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-
 CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION
 CC DOMAINS (ID1 AND ID2).
 CC -!- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED
 CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND
 CC SUFFICIENT TO PERMIT BINDING TO UNLIGAND TR AND RARs. SEQUENCES
 CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR
 CC SPECIFICITY.
 CC -!- SIMILARITY: CONTAINS 1 SANT DOMAIN (COMPOSED OF SANT-A-SANT-B).
 CC -!- SIMILARITY: CONTAINS 2 CORNR BOX.
 CC -!- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF113001; AAD20944.1; -;
 DR EMBL; AF113002; AAD20945.1; -;
 DR EMBL; AF125671; AAD22972.1; -;
 DR MGI; MGI:1337080; Ncor2.
 DR InterPro; IPR001005; Myb_DNA_bind.
 DR Pfam; PF00249; myb_DNA-binding; 2.
 DR SMART; SM00395; SANT; 2.
 DR PROSITE; PS00090; MYB_3; 1.
 KW Nuclear protein; Transcription regulation; DNA-binding; Repressor;
 KW Coiled coil; Alternative splicing.
 FT DOMAIN 165 207 COILED COIL (POTENTIAL).
 FT DOMAIN 254 312 INTERACTION WITH SIN3A/B (BY SIMILARITY).
 FT DNA_BIND 429 474 SANT-A (POTENTIAL).
 FT DNA_BIND 608 653 SANT-B (POTENTIAL).
 FT DOMAIN 492 560 COILED COIL (POTENTIAL).
 FT DOMAIN 652 682 COILED COIL (POTENTIAL).
 FT DOMAIN 775 804 PRO-RICH.
 FT DOMAIN 989 999 PRO-RICH.
 FT DOMAIN 1351 1357 PRO-RICH.
 FT DOMAIN 2094 2098 CORNR BOX OF ID1.
 FT DOMAIN 2296 2300 CORNR BOX OF ID2.
 FT DOMAIN 494 507 POLY-GLN.

FT	DOMAIN	1615	1619	POLY-ALA.
FT	DOMAIN	2434	2437	POLY-PRO.
FT	VARSPLIC	36	254	MISSING (IN ISOFORM BETA).
FT	CONFLICT	176	176	M -> RL (IN REF. 2).
FT	CONFLICT	396	402	PPMLYDA -> RHVRR (IN REF. 2).
FT	CONFLICT	555	555	D -> H (IN REF. 1; AAD20944).
FT	CONFLICT	756	756	T -> M (IN REF. 1; AAD20944).
FT	CONFLICT	785	785	V -> A (IN REF. 2).
FT	CONFLICT	806	846	HHLPHRLLTWTKKPKRLLOLPROMPRSRSLRPRRSMWE -> PSPAAPPATVDKFEQEAAPAPQOTDEAKEQKSEAEI
FT				DVG (IN REF. 2).
FT	CONFLICT	856	856	E -> K (IN REF. 1; AAD20945).
FT	CONFLICT	859	859	E -> K (IN REF. 1; AAD20945).
FT	CONFLICT	867	867	E -> K (IN REF. 1; AAD20945).
FT	CONFLICT	895	895	E -> K (IN REF. 1; AAD20945).
FT	CONFLICT	916	916	S -> F (IN REF. 1; AAD20944).
FT	CONFLICT	975	975	I -> IO (IN REF. 1; AAD20944).
FT	CONFLICT	1046	1063	PAKTPPPRWSGLPPPI -> QSTRLSPHAGHRLPSSH (IN REF. 2).
FT	CONFLICT	1073	1080	PHADPSA -> TRADPL (IN REF. 2).
FT	CONFLICT	1133	1133	MISSING (IN REF. 2).
FT	CONFLICT	1149	1149	MISSING (IN REF. 2).
FT	CONFLICT	1157	1157	G -> E (IN REF. 2).
FT	CONFLICT	1172	1201	GSATSGTUKGLPSTRAADGSPYRGLTHG -> APPPVEA SPRASQFOCRPQRLYHPR (IN REF. 2).
FT	CONFLICT	1696	1696	A -> S (IN REF. 2).
FT	CONFLICT	1857	1857	MISSING (IN REF. 2).
FT	CONFLICT	1909	1909	A -> P (IN REF. 2).
FT	CONFLICT	1913	1913	A -> G (IN REF. 2).
FT	CONFLICT	1923	1923	G -> A (IN REF. 2).
FT	CONFLICT	1956	1956	N -> S (IN REF. 2).
FT	CONFLICT	1968	1968	A -> G (IN REF. 2).
FT	CONFLICT	2195	2196	TA -> AV (IN REF. 2).
FT	CONFLICT	2213	2214	LE -> SK (IN REF. 2).
FT	CONFLICT	2224	2224	T -> A (IN REF. 2).
SQ	SEQUENCE	2472 AA:	270856 MW;	2A58F4DFB79285B CRC64;

Query Match 5.8%; Score 89; DB 1; Length 2472;

Best Local Similarity 25.3%; Pred. No. 35;

Matches 39; Conservative 20; Mismatches 53; Indels 42; Gaps 8;

QY	157	GRGAELAAIPDLDPVKRLUGLENDERAYSPAELLICEATGTPLVFDAAHHVVDK--	214
Db	765	GTALPAATQPPVPPP-----EPPAVAPAEPSVPDASGPPSP-EPShLPHRL 814	
QY	215	-----LPQEDPSVREWVLR-ARATWQPEWQVW-----HLSNGTEG 250	
Db	815	WTRNMKKPKLLQLPQKMPRSRS--LRPRSMWEKPEEPEASEEPESVKSDDKETE 872	
QY	251	PQDRRHSHLIADFPAYADVPQIEVEAKGKEAI 284	
Db	873	PEDKARG---TEAETVSEAP-LKVEEAGSKAAV 902	

RESULT 7
 P4H2_MOUSE
 ID P4H2_MOUSE STANDARD; PRT; 537 AA.
 AC Q60716;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE POLYL 4-HYDROXYLASE ALPHA-2 SUBUNIT PRECURSOR (EC 1.14.11.2).
 GN P4H2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95273376; PubMed=7753822;
 RA HeLaogski T., Annunen P., Vuori K., Macneil I.A., Pihlajaniemi T.,
 RA Kivirikko K.I.;


```
Db 422 LAESLGHVLEHGDAGRYVLTSLIDEIHE-----NDDRIAVV 459
QY 122 -----RESSVRAMSAHARVMDG----LGLARTPNMLLLHGGKGGGAEIAALIPDLPDP 172
Db 460 IDWHRVSDSRQAALGFLDONGCHHLQIIVTSWSRAGLPVGRURIGDEIAEI-----DS 514
QY 173 VRLRLGLENDERA-----YSPAELLPICEAT----- 198
Db 515 AALR--FDTDEAALLNDAGGLRLPRADVQALTTSTDGMAAALRLAALSRLGGGDATQLL 572
QY 199 -CTPLVFDAAHHVHVDKLPDQEDPSVREWLRA-----RATWQPP 237
Db 573 RGLSGASDVIIHEFLSENVDLTLEPELREFLIVASVTERTCGLASALAGITNGRAMLEEA 632
QY 238 EMOVYHLSNGIEGPDQRRHSHLIADF 263
Db 633 EHRGLFLORTEDDPNWFREHQMFADF 658

RESULT 9
APPL_MOUSE
ID APPL_MOUSE STANDARD; PRT; 653 AA.
AC Q03157;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE AMYLOID-LIKE PROTEIN 1 PRECURSOR (APLP).
GN APLP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93066322; PubMed=1279693;
RA Wasco W., Bupp K., Magendanz M., Gusella J.F., Tanzi R.E.,
RA Solomon F.;
RT "Identification of a mouse brain cDNA that encodes a protein related
RT to the Alzheimer disease-associated amyloid beta protein precursor.";
RT Proc. Natl. Acad. Sci. U.S.A. 89:10758-10762(1992).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. MAY BE PROCESSED
CC IN THE GOLGI COMPLEX; HOWEVER, IT IS NOT YET CLEAR WHETHER APLP
CC IS SECRETED.
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L04538; AAA37247.1; .
CC PIR; A46362; A46362.
CC MGD; MGI:88046; Aplp1.
CC InterPro; IPR001868; A4_APP.
CC Pfam; PF02177; A4_EXTRA; 1.
CC PRINTS; PR0203; AMYLOIDA4.
CC SMART; SM00006; A4_EXTRA; 1.
CC PROSITE; PS00319; A4_EXTRA; 1.
CC PROSITE; PS00320; A4_INTRA; 1.
CC Glycoprotein; Transmembrane; Signal.
CC SIGNAL 1 20 POTENTIAL.
CC CHAIN 21 653
CC DOMAIN 21 583 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 584 606 POTENTIAL.
CC DOMAIN 607 653 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 643 646 CLATHRIN-BINDING (POTENTIAL).
CC DOMAIN 263 271 POLY-GLU.
CC CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 584 554 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 653 AA; 72751 MW; 56516DC3EA40E4B0 CRC64;
Query Match 5.6%; Score 85.5; DB 1; Length 653;
Best Local Similarity 21.2%; Pred. No. 13;
Matches 55; Conservative 22; Mismatches 83; Indels 99; Gaps 10;
QY 53 CAADIRLRYLRLSSLPMLDLA-----GDDTGAAYLTH--LAPQLLEAG 94
Db 83 CLLDPQVRLVLEYCHQMPGLHARVEQAAQAIPMERWCGGTRSGRCAHPHHEVVPFHCLPG 142
QY 95 HATFDAGVRLLMPEQFTVLNSDRPEVRSSVR-----AMSAHARVMDGLGLARTPN 147
Db 143 EYFSEA-----LLVPEGRFLHQERMDQCESTRKHQEAQACSSQGLLIHSG----- 191
QY 148 LLLHGGKGGKGAELAAALIPDIPDPVRLRLGLENDERAYSPELLPICEATGTPLVFDAH 207
Db 192 MLLPGSDRFRGVEYVCCPP-----PATPNPSGMAAG----- 223
QY 208 HHVHDKLPDQEDPSVREWLRAATWQPPQVWVHLSNGIEGPDQRRHSHLIADFPSPA- 266
Db 224 -----DPSRSPPLGGA-----EGGEDEEE---VESFPQPV 252
QY 267 ---YADVPOIEVEAKGKEE 282
Db 253 DDYFVEPPQAEIEEEEE 271

RESULT 10
HEYL_CANFA
ID HEYL_CANFA STANDARD; PRT; 304 AA.
AC Q0TS22;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HAIRY/ENHANCER-OF-SPLIT RELATED WITH YRPV MOTIF 1 (HAIRY AND ENHANCER
DE OF SPLIT RELATED-1) (HESR-1).
GN HEYL OR HESR1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RA Pichon B., Mercan D., Pouillon V., Christophe-Hobertus C.,
RA Christophe D.;
RT "Functional cloning of nuclear proteins and nuclear targeting
RT sequences.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. "HAIRY" SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ388551; CAB65543.1; .
CC InterPro; IPR003015; HLH_Myc.
CC InterPro; IPR001092; HLH_dim.
CC Pfam; PF00010; HLH; 1.
CC SMART; SM00353; HLH; 1.
CC SMART; SM00511; ORANGE; 1.
CC PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
CC Transcription regulation; DNA-binding; Nuclear protein.
KW
```

```
FT DNA_BIND 50 62 BASIC DOMAIN.
FT DOMAIN 63 105 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 294 297 YPWM MOTIF (REQUIRED FOR ACTIVITY)
FT (POTENTIAL).
SQ SEQUENCE 304 AA: 32471 MW: 191D5BD42970F412 CRC64:

Query Match 5.5%; Score 84.5; DB 1; Length 304;
Best Local Similarity 24.7%; Pred. No. 6.1;
Matches 58; Conservative 23; Mismatches 79; Indels 75; Gaps 13:

QY 19 TISRYALPAERRE-----AKLLDLYSSNIXTLR--GAADYCAAHDIRL-YRLSSS 66
DB 68 SLSELRVPSAFKSGSAKLEAEILQTVDBLKMHTAGGKGYFDAHALMDYR---- 123
QY 67 LFPMLDAGDGTGAAYVTLHAPOLLAEAGHAFTDAGVRLMHPEQFTVLNSDRPREVRESSV 126
DB 124 -----SLGFRCLAEARVLS--IEGLDASDPLRVRLVSH-----LNNYASQ-REA-- 167
QY 127 RAMSAHARVMDGLGLARTPNW-----LLLLHGKGGRG-----AELA 163
DB 168 -ASGAHA-----GLGHLPGWSAFGHHPHVAHPLLLPQSGHGNLTGTSAPTDPHHQGRLLA 220
QY 164 ALIPDLP-----DPVRLRLGLENDERAYSPAELLPICEATGTPLVFDAHH 208
DB 221 AAHPEAPALRAPPSGGLGPV---LPVVTASAKLSPLLSSVASLSAFPFSGFSFH 272

RESULT 11
REFX5_HUMAN STANDARD; PRT; 616 AA.
ID REFX5_HUMAN
AC P48382;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-BINDING PROTEIN RFX5 (REGULATORY FACTOR X SUBUNIT 5).
GN RFX5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95262896; PubMed=7744245;
RA Steinle V., Durand B., Barras E., Zufferey M., Hadam M.R.,
RA Mach B., Reith W.;
RT "A novel DNA-binding regulatory factor is mutated in primary MHC
RT class II deficiency (bare lymphocyte syndrome).";
RL Genes Dev. 9:1021-1032(1995).
RN [2]
RP PARTIAL SEQUENCE.
RC TISSUE-Lymphoblast;
RX MEDLINE=99170284; PubMed=10072068;
RA Nagarajan U.M., Louis-Plence P., Desandro A., Nilsen R., Bushey A.,
RA Boss J.M.;
RT "RFX-B is the gene responsible for the most common cause of the bare
RT lymphocyte syndrome, an MHC class II immunodeficiency.";
RL Immunity 10:153-162(1999).
RN [3]
RP ERRATUM.
RA Nagarajan U.M., Louis-Plence P., Desandro A., Nilsen R., Bushey A.,
RA Boss J.M.;
RT Immunity 10:399-399(1999).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=2042030; PubMed=10779326;
RA Villard J., Peretti M., Masternak K., Barras E., Caretti G.,
RA Mantovani R., Reith W.;
RT "A functionally essential domain of RFX5 mediates activation of major
RT histocompatibility complex class II promoters by promoting
RT cooperative binding between RFX and NF-Y.";
RL Mol. Cell. Biol. 20:3364-3376(2000).
RN [5]
RP FUNCTION: ACTIVATES TRANSCRIPTION FROM CLASS II MHC PROMOTERS.
CC
```

```
CC RECOGNIZES X-BOXES. MEDIATES COOPERATIVE BINDING BETWEEN RFX AND
CC NF-Y. RFX BINDS THE X1 BOX OF MHC-II PROMOTERS.
CC -1- SUBUNIT: RFX CONSISTS OF AT LEAST THREE DIFFERENT SUBUNITS: RFXAP,
CC RFX5 AND RFX-B/RFXANK; WITH EACH SUBUNIT REPRESENTING A SEPARATE
CC COMPLEMENTATION GROUP. RFX FORMS COOPERATIVE DNA BINDING COMPLEXES
CC WITH X2BP AND CBF/NF-Y. RFX ASSOCIATES WITH CIITA TO FORM AN
CC ACTIVE TRANSCRIPTIONAL COMPLEX.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
CC -1- DOMAIN: THE N-TERMINUS IS REQUIRED FOR ITS ASSOCIATION WITH RFXANK
CC AND RFXAP, FOR ASSEMBLY OF THE RFX COMPLEX, AND FOR BINDING OF
CC THIS COMPLEX TO ITS X BOX TARGET SITE IN THE MHC-II PROMOTER. THE
CC C-TERMINUS MEDIATES COOPERATIVE BINDING BETWEEN THE RFX COMPLEX
CC AND NF-Y.
CC -1- PTM: PHOSPHORYLATED.
CC -1- DISEASE: DEFECTS IN RFX5 ARE A CAUSE OF HEREDITARY MHC CLASS II
CC DEFICIENCY (ALSO KNOWN AS BARE LYMPHOCYTE SYNDROME (BLS) OR HLA
CC CLASS II-DEFICIENT COMBINED IMMUNODEFICIENCY); A FORM OF SEVERE
CC COMBINED IMMUNODEFICIENCY DISEASE (SCID). RFX5 IS LINKED WITH
CC BLS COMPLEMENTATION GROUPS C.
CC -1- SIMILARITY: BELONGS TO THE RFX FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X85786; CAA59771.1; -
CC TRANSFAC: T01672; -
CC MIM: 601863; -
CC MIM: 209920; -
CC InterPro: IPR003150; RFX_DNA_binding.
CC Pfam: PF02257; RFX_DNA_binding; 1.
CC DNA-binding; Transcription regulation; Activator; Nuclear protein;
CC SCID; Phosphorylation.
CC FT DNA_BIND 92 168 POTENTIAL.
CC SEQUENCE 616 AA: 65322 MW: 5EBB33C677BB717F CRC64:
CC -----
Query Match 5.5%; Score 84.5; DB 1; Length 616;
Best Local Similarity 33.0%; Pred. No. 15;
Matches 33; Conservative 14; Mismatches 48; Indels 5; Gaps 3;

QY 47 RGAADYCAAHDIRLYRLSSLPMLDLAQDDT---GAAVLTHLAPOLLEAGHAFT-DAGV 102
DB 154 RGSKYCYSGIRKRTILVSMPLPLGLDLKSGSEPMGPEVTPAPRDELVEAFACALTCDAE 213
QY 103 RLLMHPEQFVLNSDRPREVRESSVRAMSAHARVMDGLGIA 142
DB 214 RLKRSFSSIV-EVAREFLLOQHLISARSAHIVLAKMGLA 252
RESULT 12
RPSD_CLOAB STANDARD; PRT; 378 AA.
ID RPSD_CLOAB
AC P33656;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RNA POLYMERASE SIGMA FACTOR RPOD (SIGMA-A) (MAJOR VEGETATIVE SIGMA
DE FACTOR).
GN RPOD OR SIGA OR CAC1300.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RC MEDLINE=95050216; PubMed=7961408;
CC
```

```

RA Sauer U., Treuner A., Buchholz M., Santangelo J.D., Durre P.:
RT "Sporulation and primary sigma factor homologous genes in Clostridium
RL acetobutylicum."
RN J. Bacteriol. 176:6572-6582(1994).
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
CC THEN IS RELEASED. THIS IS THE PRIMARY SIGMA-FACTOR OF THIS
CC BACTERIA.
CC -!- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z23080; CAA80625.1; -
CC EMBL; AE007642; AAK79271.1; -
CC PIR; S34307; S34307.
CC HSSP; P00579; 1STG.
CC InterPro; IPR000943; Sigma_70.
CC Pfam; PF00140; sigma70; 1.
CC PRINTS; PR00046; SIGMA70FCT.
CC PROSITE; PS00715; SIGMA70.1; 1.
CC PROSITE; PS00716; SIGMA70.2; 1.
CC Transcription regulation; Sigma factor; DNA-directed RNA polymerase;
KW DNA-binding.
FT DOMAIN 169 182 POLYMERASE CORE BINDING (POTENTIAL).
FT DNA_BIND 339 358 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 378 AA; 43260 MW; 13821F827E5D50B8 CRC64;

Query Match 5.5%; Score 84; DB 1; Length 378;
Best Local Similarity 22.5%; Pred. No. 8.8;
Matches 62; Conservative 45; Mismatches 95; Indels 74; Gaps 14;

QY 58 IRLYRLSSLPFMI-----DLA-----GD-----DTGAAVLFLHAPLLEAGHAFT 98
DB 109 VMYLKETGKVPLLSPPEEIDLAQRIKNGDRSARKKLAEANLRVSVIAKRYVGRGMLFL 168
QY 99 D-----AGVRLLMHPPOFTVLSNDRPE-----VRESSVRAMSAHARVMDGLGARTPWNL 148
DB 169 DLIOEGNGLLIKAVEKDFKGFSTYATWVHQAITRAIAQARTI-----RIPVHM 222
QY 149 -----LLHGGKGGGGAELAAALPDLP-DPVRRLRLGLENDRAYSPAEL-LP 193
DB 223 VETINKLIRVSQQLQELGRPOPEEIAKIM-DMPVDKVR-----EIMKTAQEPVSLETP 276
QY 194 ICEATGTPLVFDAHVVHVDKLPDOEDPS-----VREWVLRARATWQPPWQVHVL 244
DB 277 IGEEE-----DSH-----LGFDFIPDEAPADAAAFMLKEQLLKILNTLTTPREEKVLRL 327
QY 245 SNGIEGPDRRHSHLIADFTFSAYADVPOIEVEAKGK 280
DB 328 RFLGDDGRARTLEEVGFEFNVNTRIRIQIEAKALRK 363

RESULT 13
TRPA_MYCIT
ID TRPA_MYCIT STANDARD; PRT; 271 AA.

```

```

AC 068906;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRYPTOPHAN SYNTHASE ALPHA CHAIN (EC 4.2.1.20).
GN TRPA.
OS Mycobacterium intracellulare.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1767;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BATTY;
RA Alavi M.R., Rouse D.A., Morris S.L.;
RT "Nucleotide sequence and functional analysis of the tryptophan
RT synthase genes of Mycobacterium intracellulare";
RL Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: THE ALPHA SUBUNIT IS RESPONSIBLE FOR THE ALDOL CLEAVAGE
CC OF INDOLGLYCEROL PHOSPHATE TO INDOL AND GLYCERALDEHYDE 3-
CC PHOSPHATE.
CC -!- CATALYTIC ACTIVITY: L-SERINE + L-(INDOL-3-YL)GLYCEROL 3-PHOSPHATE
CC = L-TRYPTOPHAN + GLYCERALDEHYDE 3-PHOSPHATE + H(2)O.
CC -!- PATHWAY: LAST (FIFTH) STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
CC -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TRPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF057042; AAC17135.1; -
CC InterPro; IPR003009; FMN enzyme.
CC InterPro; IPR002028; TRP_synthase_alpha.
CC Pfam; PF00290; trp_synta; 1.
CC ProDom; PD001335; TRP_synthase_alpha; 1.
CC PROSITE; PS00167; TRP_SYNTHASE_ALPHA; 1.
KW Tryptophan biosynthesis; Lyase.
SQ SEQUENCE 271 AA; 27804 MW; BAE0C5BC0788318 CRC64;

```

```

Query Match 5.4%; Score 82.5; DB 1; Length 271;
Best Local Similarity 21.9%; Pred. No. 7.7;
Matches 32; Conservative 19; Mismatches 40; Indels 55; Gaps 4;

QY 64 SSSLPMLDLAGDDTGAAYVITHLAP-----QLLEAGHAFTDAGVR 103
DB 8 ASRLGPIFDVCRDDGGRALIGLYPTGYDPVTVSVIAMVALVESGCDIVEGVPSDPM- 66
QY 104 LLHMQEQFVLNSDRPEVRESSVRAMSAHARVMD-----GLGLARTPWNLLLL 151
DB 67 -----DGFIVKATEAALHGVGRVDRDTLAAVEALSAAGHAVVNTVNPVLR 113
QY 152 HG-----GKGGRGAEALALIPD 168
DB 114 YGIDAFARLASAGGYGLITPDLPDLP 139

RESULT 14
APPL_HUMAN
ID APPL_HUMAN STANDARD; PRT; 650 AA.
AC P51693; O00113;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE AMYLOID-LIKE PROTEIN 1 PRECURSOR (APLP).
GN APLP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```


RT "Accessory protein function in the DNA polymerase III holoenzyme from
RT E. coli.";

RL Bioessays 14:105-111(1992).

RM [8]

RP MEDLINE=93387658; PubMed=8375647;

RX Filjakowska I.J., Schaaper R.M.;

RA "Antimutator mutations in the alpha subunit of Escherichia coli DNA

polymerase III: identification of the responsible mutations and

alignment with other DNA polymerases.";

RL Genetics 134:1039-1044(1993).

CC -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME

RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.

CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.

CC THE ALPHA CHAIN IS THE DNA POLYMERASE.

CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =

CC N PYROPHOSPHATE + DNA(N).

CC -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA

CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE

DIMERIZATION TO FORM THE POLII' COMPLEX. POLIII' ASSOCIATES WITH

THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,

AND CHI) AND WITH THE BETA CHAIN. THE FINAL COMPOSITION OF THE

COMPLEX IS: (ALPHA,EPSILON,THETA)[2]-TAU[2]-(GAMMA,DELTA,DELTA',

PSI,CHI)[2]-BETA[4].

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-C FAMILY. DNAE

SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M19334; AAC36920.1; -

DR EMBL; AE000127; AAC73295.1; -

DR EMBL; D83536; BAA77859.1; -

DR EMBL; U70214; AAB08613.1; -

DR EMBL; S52931; AAB24889.1; -

DR EMBL; M96394; AAA70369.1; ALT_SEQ.

DR EMBL; D49445; BAA08424.1; -

DR PIR; C28390; DJEC3A.

DR PIR; A40637; A40637.

DR EcoGene; EG10238; dnaE.

DR InterPro; IPR003141; Pfp.N.

DR InterPro; IPR002309; trna-synt_2.

DR Pfam; PF02231; Pfp.N; 1.

DR Pfam; PF01336; trna-anti; 1.

DR SMART; SM00481; POLIIIAC; 1.

DR Transferrase; DNA-directed DNA polymerase; DNA replication;

KW Complete proteome.

SQ SEQUENCE 1160 AA; 129904 MW; 1A4F75F373841716 CRC64;

Query Match 5.43; Score 82.5; DB 1; Length 1160;
Best Local Similarity 22.5%; Pred. No. 48;
Matches 60; Conservative 30; Mismatches 84; Indels 93; Gaps 14;

QY 39 YSSNIKTLRGAADYCAAHDIRLYRLSSSLFPMDLADGDTCAAVLTHLAPOLLEAGHAFT 98

Db 58 HGAGIKPIVG-ADENVOCDL---LGDDELTHLTVLAANTGYQNLTLISKAYQRCY--- 109

QY 99 DAGVRLLMHPQFIVLNSDRPEVRESSVRAHARVWDGLGLARTPNWLLLLHGGK-GG 157

Db 110 -----GAAGPIIDRDWLITELNEGLILLGGGRMGD 138

QY 158 RGAEL---AALIPD-----LPDPVRL---RLGLENDERAYSPAELLPICEATCTPL 202

Db 139 VGRSLIRGNSALVDECAVFEHFPDRYFLELITG-RPDEESYLHA-AVELAEARGLPV 196

QY 203 V-----FDAH--HHVHVD--KLDPQEDPSVREWVLRARATWOPPEQVVHLSN 246

Db 197 VATNDVRFIDSSDFDAHEIRVAIHGFTLDDPKRP-----RNYSPQQYM----- 240

QY 247 GLEGPDQERRHSHLIADFPESAYADVPOI 273

Db 241 ----RSEEMCELFADIPEALANTVEI 263

Search completed: January 15, 2002, 14:07:28

Job time: 957 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:51:25 ; Search time 57.72 Seconds
(without alignments)
114.622 Million cell updates/sec

Title: US-09-724-296-39

Perfect score: 1527

Sequence: 1 QLGLVCLTVGPVRETRVT.....VEAKGKEAIAALRLMAPFK 294

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	94.5	6.2	270	5	PCT-US93-03035-2
2	90	5.9	1421	3	US-09-335-409-2
3	88	5.8	537	2	US-08-633-879C-2
4	85.5	5.6	634	1	US-08-339-152A-17
5	85.5	5.6	653	1	US-08-339-152A-16
6	85.5	5.6	653	2	US-08-007-999B-3
7	85.5	5.6	653	2	US-08-689-276A-3
8	84	5.5	403	4	US-08-861-774E-27
9	83	5.4	1891	2	US-08-804-227C-12
10	83	5.4	1891	2	US-08-804-198-6
11	82.5	5.4	3567	2	US-07-642-734C-4
12	82.5	5.4	3567	3	US-08-439-009A-4
13	81.5	5.3	5588	4	US-09-036-987A-6
14	81.5	5.3	5588	4	US-09-370-700-6
15	81	5.3	607	2	US-08-472-534-5
16	81	5.3	5215	4	US-09-105-537-2
17	80.5	5.3	394	6	5290690-2
18	80	5.2	841	4	US-09-413-814-107
19	80	5.2	7257	3	US-09-335-409-5
20	79.5	5.2	2004	1	US-08-375-709-15
21	79.5	5.2	2004	1	US-08-752-929-15
22	79.5	5.2	2004	4	US-09-090-793-9
23	79	5.2	1841	2	US-08-804-227C-6
24	79	5.2	3816	4	US-09-428-517-3
25	78.5	5.1	394	1	US-07-637-870-1
26	78.5	5.1	394	1	US-07-637-399-1
27	78.5	5.1	394	1	US-07-640-476-5

28	78.5	5.1	394	1	US-08-112-703-1	Sequence 1, Appli
29	78.5	5.1	3491	2	US-07-642-734C-2	Sequence 2, Appli
30	78.5	5.1	3491	3	US-08-439-009A-2	Sequence 2, Appli
31	78	5.1	3519	4	US-09-428-517-4	Sequence 4, Appli
32	77	5.0	587	4	US-08-931-608A-4	Sequence 4, Appli
33	77	5.0	719	4	US-08-975-762-59	Sequence 59, Appl
34	77	5.0	719	4	US-09-295-028-59	Sequence 59, Appl
35	77	5.0	719	4	US-09-106-582-59	Sequence 59, Appl
36	77	5.0	2890	4	US-09-413-814-67	Sequence 67, Appl
37	77	5.0	3798	3	US-09-335-409-6	Sequence 6, Appli
38	77	5.0	4472	2	US-08-804-227C-2	Sequence 2, Appli
39	76	5.0	1088	4	US-09-082-059-2	Sequence 97, Appl
40	76	5.0	1135	2	US-08-469-537A-97	Sequence 3, Appli
41	76	5.0	1864	2	US-08-804-227C-3	Sequence 42, Appl
42	76	5.0	2539	4	US-09-413-814-42	Sequence 8, Appli
43	76	5.0	4550	2	US-08-804-227C-8	Sequence 2, Appli
44	76	5.0	4550	2	US-08-804-198-2	Sequence 2, Appli
45	75.5	4.9	2289	3	US-09-051-019-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
PCT-US93-03035-2
; Sequence 2, Application PC/TUS9303035
; GENERAL INFORMATION:
; APPLICANT: ABBOTT LABORATORIES
; TITLE OF INVENTION: PURIFIED THERMOSTABLE ENDONUCLEASE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: D-377 AP0D, ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03035
; FILING DATE: 19930330
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/869,306
; FILING DATE: 16-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/860,702
; FILING DATE: 31-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BRAINARD, THOMAS D
; REGISTRATION NUMBER: 32,459
; REFERENCE/DOCKET NUMBER: 5145.PC.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-4884
; TELEFAX: 708-937-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 270 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-03035-2

Query Match 6.2%; Score 94.5; DB 5; Length 270;
Best Local Similarity 27.5%; Pred. No. 0.017;
Matches 70; Conservative 29; Mismatches 67; Indels 89; Gaps 17;

DD 370 GENERAL INQUIRY FORM (Rev. 1-6-60)

Db 167 S - - - VDDCFGLGRSAYNEGDYYHTVLWMEQVLKQ

Db 167 S - - - VDDCFGLGRSAYNEGDYYHTVLWMEQVLKQ

Db 167 S - - - VDDCFGLGRSAYNEGDYYHTVLWMEQVLKQ

QY 158 ---RGAELAAIPDLPDVRRLG-----LENDERAYS-----P 188
DB 224 DLHRAVELTRLLSL-DFSHERAGNLYRFFERLLEERGSLSNOTDAGLATQENLYRP 282
QY 189 AELLP-----ICEATGTPL-----VFDAHH-----HVV- 211
DB 283 TDYLPEDVYSLCGEGVKLTPRQKLCRYHHGNRVPLLIAPTKEEDWDSPHIVR 342
QY 212 -HDKLPDOEDSVREWV--LBARATWQPPWQV-----VHLSNGIEGQD-----R 254
DB 343 YDVMSEDEIERIKAIKPKIARATVDPKGTGVLTVASRYRVSWSLWLEDDDPVVARVNR 402
QY 255 RHSL-----IADPPSAVADVPQIEVEAKKEEAIAAL 287
DB 403 RMOHTGLTVKTAELLOVANYGMGQIEPHDFRSDDDEDAFKEL 447

RESULT 4
US-08-339-152A-17
: Sequence 17, Application US/08339152A
: Patent No. 5643726
: GENERAL INFORMATION:
: APPLICANT: Tanzi, Rudolph E.
: APPLICANT: Kovacs, Dora M.
: TITLE OF INVENTION: Methods For Modulating Transcription
: TITLE OF INVENTION: From The Amyloid -Protein Precursor (APP) Promoter
: NUMBER OF SEQUENCES: 33
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
: STREET: 1100 New York Ave., NW, Suite 600
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/339,152A
: FILING DATE: 10-NOV-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Steffe, Eric K.
: REGISTRATION NUMBER: 36,688
: REFERENCE/DOCKET NUMBER: 0609.4120000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-371-2600
: TELEFAX: 202-371-2540
: TELEX:
: INFORMATION FOR SEQ ID NO: 17:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 634 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
US-08-339-152A-17

Query Match 5.6%; Score 85.5; DB 1; Length 634;
Best Local Similarity 21.28; Pred. No. 0.69;
Matches 55; Conservative 22; Mismatches 83; Indels 99; Gaps 10;

QY 53 CAAHDIRLYRLSSSLFPLDLA-----GDDTGAVALTH--LAPQLLEAG 94
DB 63 CLLDQVRVLEYCRQMPYELHIAVEQAQAIPMERWCGGTRSGRCAHPHHEVVPFHCLPG 122
QY 95 HAFTDAGVRLMLHPQFIVLNSDRPREVRESSVR-----AMSAHARVMDGLGARTPNW 147
DB 123 EFPSEA----LLVPEGCRFLHQERMDQCESSTRRHQEAQACSSQGLILHSG----- 171
QY 148 LLLHGGKGRGAELAAIIPDLDPVRLRLGLENDERAYSAPALLPICEATGTPLVFDH 207

DB 172 MLLPCGSDRFRGVEYVCCPP-----PATPNPSGMAAG----- 203
QY 208 HHVVHDKLPDOEDSPSVREWVLRARATWQPPWQVHLSNGIEGQDPRRHSHLIADFPSPA- 266
DB 204 -----DPSTRSWPLGGRA-----EGGEDEEE---VESFPQPV 232
QY 267 ---YADVPOJFEVAKKEE 282
DB 233 DDFVEPFA-----EEEEEEEE 251

RESULT 5
US-08-339-152A-16
: Sequence 16, Application US/08339152A
: Patent No. 5643726
: GENERAL INFORMATION:
: APPLICANT: Tanzi, Rudolph E.
: APPLICANT: Kovacs, Dora M.
: TITLE OF INVENTION: Methods For Modulating Transcription
: TITLE OF INVENTION: From The Amyloid -Protein Precursor (APP) Promoter
: NUMBER OF SEQUENCES: 33
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
: STREET: 1100 New York Ave., NW, Suite 600
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/339,152A
: FILING DATE: 10-NOV-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Steffe, Eric K.
: REGISTRATION NUMBER: 36,688
: REFERENCE/DOCKET NUMBER: 0609.4120000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-371-2600
: TELEFAX: 202-371-2540
: TELEX:
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 653 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-339-152A-16

Query Match 5.6%; Score 85.5; DB 1; Length 653;
Best Local Similarity 21.28; Pred. No. 0.73;
Matches 55; Conservative 22; Mismatches 83; Indels 99; Gaps 10;

QY 53 CAAHDIRLYRLSSSLFPLDLA-----GDDTGAVALTH--LAPQLLEAG 94
DB 83 CLLDQVRVLEYCRQMPYELHIAVEQAQAIPMERWCGGTRSGRCAHPHHEVVPFHCLPG 142
QY 95 HAFTDAGVRLMLHPQFIVLNSDRPREVRESSVR-----AMSAHARVMDGLGARTPNW 147
DB 143 EFPSEA----LLVPEGCRFLHQERMDQCESSTRRHQEAQACSSQGLILHSG----- 191
QY 148 LLLHGGKGRGAELAAIIPDLDPVRLRLGLENDERAYSAPALLPICEATGTPLVFDH 207
DB 192 MLLPCGSDRFRGVEYVCCPP-----PATPNPSGMAAG----- 223
QY 208 HHVVHDKLPDOEDSPSVREWVLRARATWQPPWQVHLSNGIEGQDPRRHSHLIADFPSPA- 266

Db 224 -----DPSTRSWPLGGRA-----EGGEDEEE---VESFPQPV 252
QY 267 ---YADVPOIEVEAKKEE 282
Db 253 DDYFVEPPQAEDEEEEEE 271

RESULT 6

US-08-007-999B-3
; Sequence 3, Application US/08007999B
; Patent No. 5851787
; GENERAL INFORMATION:
; APPLICANT: Wasco, Wilma
; APPLICANT: Bupp, Keith
; APPLICANT: Magendantz, Margaret
; APPLICANT: Tanzi, Rudolph
; APPLICANT: Solomon, Frank
; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox, P.L.L.C.
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/007.999B
; FILING DATE: 21-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,642
; FILING DATE: 20-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,022
; FILING DATE: 17-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, G. Kevin
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: 0609.3520002/JAG/GKT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2571
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 653 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-007-999B-3

Query Match 5.6%; Score 85.5; DB 2; Length 653;
Best Local Similarity 21.2%; Pred. No. 0.73;
Matches 55; Conservative 22; Mismatches 83; Indels 99; Gaps 10;

QY 53 CAADHRLRYLSSSLFPMDLA-----GDDTGAAYLTH--LAPOLLEAG 94
Db 83 CLLDPPQVLEYCRQYMPHILHARVEQAQAIPMERWCGTGRGCAHPHVEVPFHCLPG 142
QY 95 HAFTDAGVRLMHPEQFIVLNSDRPEVRESSVR-----AMSAHARVMDGLGLARTPN 147
Db 143 EFVSEA-----LLVPEGCRFLHQRMDQCSSTRRHOEAQACSSQGLILHSG----- 191
QY 148 LLLHGGKGRGAELAAALIPDPVRLRLGLENDERAYSFAELLPICEATGTPLVFAH 207
Db 192 MLLPCGSDRFRGVVEVCCPP-----PATPNPGMAAG----- 223

QY 208 HHVHDKLPQEDPSVREWVLRAEATWQPPENQVHLSNGIEGPQDRRHSHLIADPPSA- 266
Db 224 -----DPSTRSWPLGGRA-----EGGEDEEE---VESFPQPV 252
QY 267 ---YADVPOIEVEAKKEE 282
Db 253 DDYFVEPPQAEDEEEEEE 271

RESULT 7

US-08-689-276A-3
; Sequence 3, Application US/08689276A
; Patent No. 5891691
; GENERAL INFORMATION:
; APPLICANT: Wasco, Wilma
; APPLICANT: Bupp, Keith
; APPLICANT: Magendantz, Margaret
; APPLICANT: Tanzi, Rudolph
; APPLICANT: Solomon, Frank
; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/689,276A
; FILING DATE: 06-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/007,999
; FILING DATE: 21-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,642
; FILING DATE: 20-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,022
; FILING DATE: 17-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: JORGE A. GOLDSTEIN
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 0609.3520003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2571
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 653 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-689-276A-3

Query Match 5.6%; Score 85.5; DB 2; Length 653;
Best Local Similarity 21.2%; Pred. No. 0.73;
Matches 55; Conservative 22; Mismatches 83; Indels 99; Gaps 10;

QY 53 CAADHRLRYLSSSLFPMDLA-----GDDTGAAYLTH--LAPOLLEAG 94
Db 83 CLLDPPQVLEYCRQYMPHILHARVEQAQAIPMERWCGTGRGCAHPHVEVPFHCLPG 142
QY 95 HAFTDAGVRLMHPEQFIVLNSDRPEVRESSVR-----AMSAHARVMDGLGLARTPN 147
Db 143 EFVSEA-----LLVPEGCRFLHQRMDQCSSTRRHOEAQACSSQGLILHSG----- 191

```

RESULT          9
US-08-804-227C-12
; Sequence 12, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501

```

RESULT 10
US-08-804-198-6
: Sequence 6, Application US/08804198
: Patent No. 5945320
: GENERAL INFORMATION:
: APPLICANT: Burgett, Stanley G.
: APPLICANT: Kuhstoss, Stuart A.
: APPLICANT: Rao, Nagaraja R.
: APPLICANT: Richardson, Mark A.
: APPLICANT: Rosteck, Paul R., Jr.
: TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESS: PAUL R. CANTRELL 1138
: STREET: LILLY CORPORATE CENTER

```

; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804.198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1891 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-804-198-6

```

```

Query Match      5.4%; Score 83; DB 2; Length 1891;
Best Local Similarity 22.5%; Pred. No. 7.6;
Matches 78; Conservative 28; Mismatches 126; Indels 114; Gaps 16;

QY 16 RTVLSRYRALSPAEREAKLLDLYSSNIKTLRGAADYCAAHDIRLYRLSSSLFFPMLDLG 75
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 768 RTKRLRVSHAFSPMDMLAD-----FRAVDYVDYHAPLPVVS-----EVTG 812

QY 76 DDTGAARVLTLPOLLEAGH-----AFTDAGVRLLMHPPEQFI----- 112
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 813 DLADAA-----QLTDPGYWTROVROPYRFADAVRTASARDAATFIELGPDVAVLCGMAE 865

QY 113 -----VLNSDRPEYRESSVRAHARVMDGLGLARTPWNLLLLHGGKGGK 159
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 866 ESLAERADVFPALRGPEGDTVLRAASAYVR---GAGL---DW---AALYGGTGARR 917

QY 160 AEL -----AALIPDLPD-VRLRLGLENDERAYSPAELLPICEATGTP 202
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 918 TDLPTVAFQHSRYLWAPASAAVAPATAFSPVPEAEQDGALWAAVHAGDVASAAARLG 977

QY 203 VFDAH-HHVVHDKLP-----DQEDPSVREWVLRLARATWQPEWVHLSNGIEGPQDR- 254
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 978 ADDAGTEHELRAVLPHLAWHDRDRATARTAGLHYRVTVQALTEADAVRFS-----PSDRW 1032

QY 255 -----RHSHLIADFPAYADVFPQIEVEAKGKEATAALRLMAPFK 294
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1033 LMVEHQHTEC-----ADAAERALRAAGAE-----VTRLVWPLE 1066

```

```

RESULT 11
US-07-642-734C-4
; Sequence 4, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
; APPLICANT: Katz, L.
; APPLICANT: Donadio, S.
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Recombinant DNA Method for Producing
; TITLE OF INVENTION: Erythromycin Analogs
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edward H. Gorman
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; CITY: Abbott Park

```

```

; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/642,734C
; FILING DATE: 17-JAN-91
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Danckers, Andreas M
; REGISTRATION NUMBER: 32652
; REFERENCE/DOCKET NUMBER: 4952.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-9396
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3567 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-642-734C-4

```

```

Query Match      5.4%; Score 82.5; DB 2; Length 3567;
Best Local Similarity 23.8%; Pred. No. 24;
Matches 64; Conservative 28; Mismatches 94; Indels 83; Gaps 13;

QY 18 VTLRSRYRALSP-----ADREAKLLDLYSSNIKTLRGAADYCAAHDIRLYRLSSSLFFPMLDL 73
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1847 VLAAMRGMLRSLRHADLSPLHDWESGAVEVLREEVWPAGERPRAGVSS-----FGV 1901

QY 74 AGDGTGAARVLTLPOLLEAGHAFDTAGVRLLMHPPEQFIVLNSDRPE-VRESSVRAHSAH 132
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1302 SG--TNAIVIVEAPABQEA--ARTERG-----PLPFVL--SGRSEAVVAAQARALAEH 1949

QY 133 ARVMDGLGLARTPWL-----LLHGGKGGKGAELALIPDLPDPVRLRLGLE 181
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1950 LRDTPELGLTDAWTATLATGRARFDVRAAVLGGDDRAGVCAELDALAEGRP----- 1998

QY 182 DERAYSPAELLPICEATGTP-LVFDAHHVHVHDKLPDQEDPSVREWVLRLARATWQPE-- 238
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1999 ----SADAVAPVTSAPRKPVLF-----PQQG----AQWVGWARDLLESSEVF 2038

QY 239 -----WQVVHLSNGIEGP 251
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2039 AESMSRCAEALSPHTDWKLLQVVRGDDGP 2067

```

```

RESULT 12
US-08-439-009A-4
; Sequence 4, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polypeptides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven F. Weinstein
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:

```

;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/439,009A
;; FILING DATE: 11-MAY-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Casuto, Dianne
;; REGISTRATION NUMBER: 40,943
;; REFERENCE/DOCKET NUMBER: 4952.US.D1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 847-938-3137
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3567 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-439-009A-4

Query Match 5.4%; Score 82.5; DB 3; Length 3567;
Best Local Similarity 23.8%; Pred. No. 24;
Matches 64; Conservative 28; Mismatches 94; Indels 83; Gaps 13;

Qy 18 VTLGRALSP-----AREAKLLDLYSNITKTGGAADYCAAHDIRLYRSSLPMLDL 73
Db 1847 VVLAHRHGLPRSLHADSPLHDSWEGAVEVLRPEVPWAGERPRRAGVSS-----FGV 1901
Qy 74 AGDDTGAAVLTHLAPOLLEAGHAFTDAGVRLLMHPEQFIVLNSDRPE-VRESSVRAMSAH 132
Db 1902 SG-TNAHVIVVEEAPAEQEA-ARTERG-----PLPFVL-SCRSVAVVAAQARALAEH 1949
Qy 133 ARVMDGLGLARTPNWL-----LLHGGKGGGGAELAAALIPOLPDVPRVRLGLN 181
Db 1950 LRDTPGLGLTDAWTLATGRADFVRAVGLDGRAGVCAELDLAAGERP----- 1998
Qy 182 DERAYSPAELLPICEATGTP-LVFDAAHHVVDKLPQEDPSVREWVLRARATWQPPV 238
Db 1999 -----SADAVAPVTSAPRKPVLPF-----PQGG-----AQWVGMDRLLESSEVF 2038
Qy 239 -----WQVHLSNGIEGP 251
Db 2039 AESMSRCAEALSPHTDMKLLDVVRGDDGP 2067

RESULT 13
US-09-036-987A-6
; Sequence 6, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/036,987A
;; FILING DATE: 09-MAR-1998
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Stuart, Donald R.
;; REGISTRATION NUMBER: 28,479
;; REFERENCE/DOCKET NUMBER: 50,608
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (317)337-4816
;; TELEFAX: (317)337-4847
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 5588 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-09-036-987A-6

Query Match 5.3%; Score 81.5; DB 4; Length 5588;
Best Local Similarity 21.8%; Pred. No. 64;
Matches 62; Conservative 31; Mismatches 103; Indels 89; Gaps 12;

Qy 22 RYALSPAEREAKLLDLYSNITKTLRG-----AADYCAAHDIRLYRLSS 65
Db 5204 RLAGLSEGERQQVQLQRVADIIVLGHGRSSDVDTIEKPLAELGFDLSFAIELR-NRLAT 5262
Qy 66 SLFPML--DLAGD-DTCAAVLTHLAPOLLEAGHAFTDAGVRLLMHPEQFIVLNSDRPEVR 122
Db 5263 ATGLRLPATLAFDHGTAALAOHVCAQL---GTATAPAPRR-----TDNDATPEV 5310
Qy 123 ESSVRAMSAHARVMDGLGLARTPNWL---LLHGGKGGGGAELAAALIPOLPDVPRVRLGLN 182
Db 5311 RSLFQQYAAAGRLDGMVLKVAQAQLRPVFGSPGE-----LESLEPKVQLSRG---- 5358
Qy 183 ERAYSPAELLPICEATGTP-LVFDAAHHVVDKLPQEDPSVREWVLRARATWQPPVQVV 242
Db 5359 -----PEELALVC-----MPALIG-----MPPAQQYA 5380
Qy 243 HLSNGIEGPQDRRHSHLIADFPSAYADVP---QIEVEAKGKEAI 284
Db 5381 RTAAGFRDVRDVS-----VIPMPGFIAGEPLPSAIEVAVRTQAEAV 5421

RESULT 14
US-09-370-700-6
; Sequence 6, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 5588
; TYPE: PRT
; ORGANISM: Saccharopolyspora spinosa
; US-09-370-700-6

Query Match	5.3%;	Score 81;	DB 2;	Length 607;
Best Local Similarity	21.0%;	Pred. No. 2.1;		

Search completed: January 15, 2002, 13:51:29
Job time: 358 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:07:20 ; Search time 45.07 Seconds
(without alignments)
471.835 Million cell updates/sec

Title: US-09-724-296-38
Perfect score: 3055
Sequence: 1 MGTTLGSLSLGRGAAPTV.....HKLNKYHDWLWENHEKLSLS 580

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	208.5	6.8	486	1 MEC2_HUMAN	P51608 rat mus sapien
2	201	6.6	492	1 MEC2_RAT	Q00566 rattus norv
3	198.5	6.5	484	1 MEC2_MOUSE	Q92246 mus musculus
4	145	4.7	1301	1 SAC3_YEAST	P46674 saccharomyc
5	145	4.7	3418	1 BRC2_HUMAN	P51587 homo sapien
6	138.5	4.5	500	1 GAR2_SCHPO	P41891 schizosacch
7	137	4.5	560	1 YJK5_YEAST	P42948 saccharomyc
8	136.5	4.5	598	1 CYL1_HUMAN	P35663 homo sapien
9	133	4.4	1447	1 BUD4_YEAST	P47136 saccharomyc
10	131.5	4.3	997	1 BTR1_SCHPO	O14064 schizosacch
11	122	4.0	3924	1 ANK2_HUMAN	Q01484 homo sapien
12	120.5	3.9	1147	1 AC15_HUMAN	P35251 homo sapien
13	120.5	3.9	1233	1 YF16_YEAST	P43597 saccharomyc
14	120	3.9	1131	1 AC15_MOUSE	P35601 mus musculus
15	120	3.9	1658	1 YMK7_YEAST	Q03661 saccharomyc
16	119	3.9	886	1 ORCL_KLULA	P54788 kluyveromyc
17	118	3.9	704	1 NP14_RAT	P41777 rattus norv
18	118	3.9	3329	1 BRC2_MOUSE	P97929 mus musculus
19	117	3.8	1273	1 YAR2_SCHPO	Q10135 schizosacch
20	117	3.8	1850	1 VIT2_CHICK	P02845 gallus gall
21	116.5	3.8	633	1 MLH1_TETTH	P40631 tetrahymena
22	116	3.8	1290	1 XPC_KENLA	P50532 xenopus lae
23	116	3.8	3256	1 K167_HUMAN	P46013 homo sapien
24	115.5	3.8	2156	1 ORP1_HUMAN	P56715 homo sapien
25	114.5	3.7	970	1 FSU1_YEAST	P53550 saccharomyc
26	114.5	3.7	1213	1 FMN_CHICK	Q05858 gallus gall
27	114	3.7	783	1 YAYB_SCHPO	O10218 schizosacch
28	114	3.7	805	1 TAC1_HUMAN	O75410 homo sapien
29	114	3.7	1539	1 Y373_HUMAN	O15078 homo sapien
30	113.5	3.7	1164	1 BAG_STRAG	P27951 streptococc
31	113	3.7	482	1 YSR2_CAEEL	Q09950 caenorhabdi
32	113	3.7	727	1 PEC1_MOUSE	Q08481 mus musculus
33	112.5	3.7	676	1 H57C_TRYBB	P20030 trypanosoma

ALIGNMENTS

RESULT 1
MEC2_HUMAN
ID MEC2_HUMAN STANDARD; PRT; 486 AA.
AC P51608; O15233;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE METHYL-CPG-BINDING PROTEIN 2 (MECP-2 PROTEIN) (MECP2).
GN MECP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kudo S., Fukuda M.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Thiesen J., Straetling W.H.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-97130625; PubMed-8976388;
RA Vilain A., Apiou F., Vogt N., Dutrillaux B., Malfoy B.;
RT "Assignment of the gene for methyl-CpG-binding protein 2 (MECP2) to human chromosome band Xq28 by in situ hybridization.";
RL Cytogenet. Cell Genet. 74:293-294(1996).
RN [4]
RP SEQUENCE FROM N.A.
RA Reichwald K., Rosenthal A., Kioschis P., Platzer M.;
RT "Mapping and sequence analysis of the human MECP2 locus.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 10-486 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE-96327611; PubMed-8672133;
RA D'Esposito M., Quaderi N.A., Ciccocioppa A., Bruni P., Esposito T.,
D'Esposito M., Brown S.D.M.;
RT "Isolation, physical mapping, and Northern analysis of the X-linked human gene encoding methyl CpG-binding protein, MECP2.";
RL Mamm. Genome 7:533-535(1996).
RN [6]
RP SEQUENCE OF 10-486 FROM N.A.
RA Reichwald K., Bauer D., Brenner V., Drescher B., Coy J.,
RA Kioschis P., Korn B., Nyakatura G., Platzer M., Poustka A.,
RA Sandoval N., Rosenthal A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [7]
RP VARIANTS RTT TRP-106; CYS-133; SER-155 AND MET-158.
RX MEDLINE-99438392; PubMed-10508514;
RA Amir R.E., Van den Veyver I.B., Wan M., Tran C.Q., Francke U.,
Zoghbi H.Y.;
RT "Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2.";

34 112.5 3.7 1522 1 PSTL_SCHPO
35 112 3.7 1462 1 NKCR_HUMAN
36 111.5 3.6 623 1 HPC2_YEAST
37 111.5 3.6 906 1 CENC_MOUSE
38 111 3.6 549 1 MIF2_YEAST
39 111 3.6 1293 1 XPC_DROME
40 111 3.6 1828 1 MAP2_MOUSE
41 110.5 3.6 667 1 CYL1_BOVIN
42 110.5 3.6 2014 1 YJ07_YEAST
43 110 3.6 594 1 ZF37_MOUSE
44 110 3.6 872 1 S3B2_HUMAN
45 110 3.6 1076 1 NUPL_YEAST

Q09750 schizosacch
P30414 homo sapien
Q01448 saccharomyc
P49452 mus musculus
P35201 saccharomyc
Q24595 drosophila
P20357 mus musculus
P35662 bos taurus
P39526 saccharomyc
P17141 mus musculus
Q13435 homo sapien
P20676 saccharomyc

RL Nat. Genet. 23:185-188(1999).
 RN [8]
 RP VARIANT RTT VAL-140.
 RX MEDLINE=20465115; PubMed=11007980;
 RA Orrico A., Lam C., Galli L., Dotli M.T., Hayek G., Tong S.F.,
 RA Poon P.M., Zappella M., Federico A., Sorrentino V.;
 RT "MECP2 mutation in male patients with non-specific X-linked mental
 RT retardation.";
 RL FEBS Lett. 481:285-288(2000).
 RN [9]
 RP VARIANTS RTT W-106; P-124; C-13; C-134; R-152; M-158 AND C-306.
 RX MEDLINE=20439334; PubMed=10991688;
 RA Obata K., Matsushita T., Yamashita Y., Fukuda T., Kuwajima K.,
 RA Horiuchi I., Nagamitsu S., Iwanaga R., Kimura A., Omori I., Endo S.,
 RA Mori K., Kondo I.;
 RT "Mutation analysis of the methyl-CpG binding protein 2 gene (MECP2) in
 RT patients with Rett syndrome.";
 RL J. Med. Genet. 37:608-610(2000).
 RN [10]
 RP VARIANTS RTT R-101; W-106; M-158 AND C-306, AND VARIANT K-397.
 RX MEDLINE=20439335; PubMed=10991689;
 RA Hampson K., Woods C.G., Latif F., Webb T.;
 RT "Mutations in the MECP2 gene in a cohort of girls with Rett
 RT syndrome.";
 RL J. Med. Genet. 37:610-612(2000).
 CC -!- FUNCTION: CHROMOSOMAL PROTEIN THAT BINDS TO METHYLATED DNA. IT CAN
 CC BIND SPECIFICALLY TO A SINGLE METHYL-CpG PAIR. IT IS NOT
 CC INFLUENCED BY SEQUENCES FLANKING THE METHYL-CpGS. MEDIATES
 CC TRANSCRIPTIONAL REPRESSION THROUGH INTERACTION WITH HISTONE
 CC DEACETYLASE AND THE COREPRESSOR SIN3A.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR. COLOCALIZED WITH METHYL-CpG IN THE
 CC GENOME.
 CC -!- TISSUE SPECIFICITY: PRESENT IN ALL ADULT SOMATIC TISSUES TESTED.
 CC -!- DISEASE: DEFECTS IN MECP2 ARE THE CAUSE OF RETT SYNDROME (RTT), AN
 CC X-LINKED DOMINANT DISEASE. RTT IS A PROGRESSIVE NEUROLOGIC
 CC DEVELOPMENTAL DISORDER AND ONE OF THE MOST COMMON CAUSES OF MENTAL
 CC RETARDATION IN FEMALES. PATIENTS APPEAR TO DEVELOP NORMALLY UNTIL
 CC 6 TO 18 MONTHS OF AGE, THEN GRADUALLY LOOSE SPEECH AND PURPOSEFUL
 CC HAND MOVEMENTS AND DEVELOP MICROCEPHALY, SEIZURES, AUTISM, ATAXIA,
 CC INTERMITTENT HYPERVENTILATION, AND STEREOTYPIC HAND MOVEMENTS.
 CC AFTER INITIAL REGRESSION, THE CONDITION STABILIZES AND PATIENTS
 CC USUALLY SURVIVE INTO ADULTHOOD.
 CC -!- SIMILARITY: CONTAINS (R)GRP(K) MOTIFS AND SPKK MOTIFS THAT HAVE
 CC BEEN FOUND TO BIND TO MINOR GROOVE OF AT-RICH DNA.
 CC -!- SIMILARITY: CONTAINS 1 METHYL-BINDING DOMAIN (MBD).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; L37298; AAC32737.1; -;
 DR EMBL; Y12643; CAA73190.1; -;
 DR EMBL; X99686; CAA68001.1; -;
 DR EMBL; AF030876; AAC08757.1; -;
 DR EMBL; AF031078; AAC08758.1; -;
 DR EMBL; X89430; CAA61599.1; -;
 DR EMBL; X94628; CAA64331.1; -;
 DR MIM; 300005; -;
 DR MIM; 312750; -;
 DR InterPro; IPR000637; AT hook.
 DR InterPro; IPR001739; MBD.
 DR Pfam; PF01429; MBD; 1.
 DR SMART; SM00384; AT_hook; 1.
 DR SMART; SM00391; MBD; 1.
 KW Transcription regulation; Repressor; DNA-binding; Nuclear protein;
 KW Disease mutation; Polymorphism.
 FT DOMAIN 96 149 MBD.
 FT DOMAIN 277 283 POLY-ALA.
 FT DOMAIN 366 372 POLY-HIS.

FT DOMAIN 384 393 POLY-PRO.
 FT VARIANT 101 101 P -> R (IN RTT).
 FT 106 106 /FTID=VAR_010276.
 FT 124 124 R -> W (IN RTT).
 FT 124 124 /FTID=VAR_010272.
 FT 133 133 L -> F (IN RTT).
 FT 133 133 /FTID=VAR_010277.
 FT 134 134 R -> C (IN RTT).
 FT 134 134 /FTID=VAR_010273.
 FT 140 140 S -> C (IN RTT).
 FT 140 140 /FTID=VAR_010278.
 FT 152 152 A -> V (IN RTT).
 FT 152 152 /FTID=VAR_010279.
 FT 155 155 P -> R (IN RTT).
 FT 155 155 /FTID=VAR_010280.
 FT 158 158 E -> S (IN RTT).
 FT 158 158 /FTID=VAR_010274.
 FT 201 201 T -> M (IN RTT).
 FT 201 201 /FTID=VAR_010275.
 FT 306 306 A -> V.
 FT 306 306 /FTID=VAR_010281.
 FT 397 397 R -> C (IN RTT).
 FT 397 397 /FTID=VAR_010282.
 FT 72 75 E -> K.
 FT 290 290 /FTID=VAR_010283.
 FT 486 486 PAVP -> RLC (IN REF. 5).
 FT 486 486 CONFLICT 290 290 R -> G (IN REF. 3).
 SQ SEQUENCE 486 AA; 52440 MW; EB6A33233AEDA566 CRC64;
 Query Match 6.8%; Score 208.5; DB 1; Length 486;
 Best Local Similarity 23.9%; Pred. No. 2.5e-06;
 Matches 104; Conservative 53; Mismatches 149; Indels 129; Gaps 17;
 QY 37 EDVAMELERVCEDEEOMNIKRSECNPLLOEPIASAQFG-----ATACT 80
 DB 22 KDRPLRFKVKYKDKKEGKHEPVOPSAHSAEPAEACKAETSEGGSPAPVPEASASP 81
 QY 81 ECRKSV-----PCGWERVVKQRLFOKTAGRDFDVFISPOGLKFRSKSLANY 127
 DB 82 KQRSITRDRCPMYDDPTLPEGWTRKLKQKRSRAGKYDVYLLINPOGKAFRSKVELIAY 141
 QY 128 LHRNGETSLKPEDFDFTVLSKRGIKSRKYKDCSMAALTSHLQNSNNSNNMLRTRSKCKKD 187
 DB 142 FEKVGDTSLDNDFDFTV-TGRGSPSR-----REQKEPKK- 175
 QY 188 VFMPSSSSSELQSRGL---SNETSTHLLKKEGVDVDFNFRKVRPKGVTKLGIPIK 244
 DB 176 ---PKSPKAPGTGRGRPKSGTTRPKAATSEGQVK--RVLEKSPGK--LLVKMPF- 226
 QY 245 KTKKGRKSCSGFVQSDS---KRESVCNKADAESPVAKSOLDRTVCISDAGACGETL 300
 DB 227 QTSFGGKAEGGATTTQVMVIRKGRKRAADPAIPKR-----GRKPGSVV 276
 QY 301 SVTSEENSLVKKERSLSGSCNFCSEQKTSGLINKFCSAKDEHNEKYEDTFLESEIGT 360
 DB 277 AAAAAEAKKAVKESIR-----SVQETVLPKK-----RKTRET-----V 312
 QY 361 KVEVVERKEHLHTDIL--KRGSEMDNNCSPTKDFTGCEKIFQEDTIPRTQIERKTSLYF 418
 DB 313 STIEKVEVKPLIVSTLGEKSGKGLKTKCSGPKR-----SKESPCKGR----- 354
 QY 419 SSKYNEALSPPRRK 433
 DB 355 ----SSASPPPKKE 365
 RESULT 2
 MEC2_RAT
 ID MEC2_RAT STANDARD; PRT; 492 AA.
 AC Q00566;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE BREAST CANCER TYPE 2 SUSCEPTIBILITY PROTEIN.
 GN BRCA2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96112016; PubMed=8524414;
 RA Wooster R., Bignell G., Lancaster J., Swift S., Seal S., Mangion J.,
 RA Collins N., Gregory S., Gumbs C., Micklem G., Barfoot R., Hamoudi R.,
 RA Patel S., Rice C., Biggs P., Hashim Y., Smith A., Connor F., Tonin P.,
 RA Arason A., Gudmundsson J., Fienec D., Kelsell D., Ford D., Tonin P.,
 RA Bishop D.T., Spurr N.K., Ponder B.A.J., Eeles R., Peto J., Devilee P.,
 RA Cornelisse C., Lynch H., Narod S., Lenoir G., Egilsson V.,
 RA Barkadottir R.B., Easton D.F., Bentley D.R., Futreal P.A.,
 RA Ashworth A., Stratton M.R.;
 RT "Identification of the breast cancer susceptibility gene BRCA2.";
 RL Nature 378:789-792(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96172838; PubMed=8589730;
 RA Tavtigian S.V., Simard J., Rommens J., Couch F., Shattuck-Eidens D.,
 RA Neuhausen S., Merajver S., Thorlacius S., Offit K., Stoppa-Lyonnet D.,
 RA Belanger C., Bell R., Berry S., Bogden R., Chen Q., Davis T.,
 RA Dumont M., Frye C., Hattier T., Jammulapati S., Janecki T., Jiang P.,
 RA Kehrer R., Leblanc J.-F., Mitchell J.T., McArthur-Morrison J.,
 RA Nguyen K., Peng Y., Samson C., Schroeder M., Snyder S.C., Steele L.,
 RA Stringfellow M., Stroup C., Swedlund B., Swensen J., Teng D.,
 RA Thomas A., Tran T., Tran T., Tranchant M., Weaver-Feldhaus J.,
 RA Wong A.K.C., Shiruya H., Eyfjord J.E., Cannon-Albright L., Labrie F.,
 RA Skolnick M.H., Weber B., Kamb A., Goldfar D.E.;
 RT "The complete BRCA2 gene and mutations in chromosome 13q-linked
 RT kindreds.";
 RL Nat. Genet. 12:333-337(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Hunt S., McMurray A., Williamson H.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP VARIANT OC HIS-2787, AND VARIANTS ASN-372; MET-1915 AND VAL-2466.
 RX MEDLINE=96275740; PubMed=8665505;
 RA Takahashi H., Chiu H.-C., Bandera C.A., Behbakht K., Liu P.C.,
 RA Couch F.J., Weber B.L., LiVolsi V.A., Furusato M., Rebane B.A.,
 RA Cardonick A., Benjamin I., Morgan M.A., King S.A., Mikuta J.J.,
 RA Rubin S.C., Boyd J.;
 RT "Mutations of the BRCA2 gene in ovarian carcinomas.";
 RL Cancer Res. 56:2738-2741(1996).
 RN [5]
 RP VARIANTS ASN-372; ASP-991; SER-1147; MET-1915 AND CYS-2034.
 RX MEDLINE=96241588; PubMed=8673091;
 RA Couch F.J., Farid L.M., Deshano M.L., Tavtigian S.V., Calzone K.,
 RA Campeau L., Peng Y., Bogden R., Chen Q., Neuhausen S.,
 RA Shattuck-Eidens D., Godwin A.K., Daly M., Radford D.M., Sedlacek S.,
 RA Rommens J., Simard J., Garber J., Merajver S., Weber B.L.;
 RT "BRCA2 germline mutations in male breast cancer cases and breast
 RT cancer families.";
 RL Nat. Genet. 13:123-125(1996).
 RN [6]
 RP VARIANT GLU-3095.
 RX MEDLINE=96225456; PubMed=8640235;
 RA Lancaster J.M., Wooster R., Mangion J., Phelan C.M., Cochran C.,
 RA Gumbs C., Seal S., Barfoot R., Collins N., Bignell G., Patel S.,
 RA Hamoudi R., Larsson C., Wiseman R.W., Berchuck A., Iglehart J.D.,
 RA Marks J.R., Ashworth A., Stratton M.R., Futreal P.A.;
 RT "BRCA2 mutations in primary breast and ovarian cancers.";
 RL Nat. Genet. 13:238-240(1996).
 RN [7]
 RP VARIANTS.
 RX MEDLINE=96225457; PubMed=8640236;
 RA
 CC
 CC -!- FUNCTION: MAY PARTICIPATE IN A PATHWAY ASSOCIATED WITH THE
 CC ACTIVATION OF DOUBLE-STRAND BREAK REPAIR AND/OR HOMOLOGOUS
 CC RECOMBINATION.
 RA Teng D.H.-F., Bogden R., Mitchell J., Baumgard M., Bell R., Berry S.,
 RA Davis T., Ha P.C., Kehrer R., Jammulapati S., Chen Q., Offit K.,
 RA Skolnick M.H., Tavtigian S.V., Jhanwar S., Swedlund B., Wong A.K.C.,
 RA Kamb A.;
 RT "Low incidence of BRCA2 mutations in breast carcinoma and other
 RT cancers.";
 RL Nat. Genet. 13:241-244(1996).
 RN [8]
 RP VARIANT ASN-2415.
 RX MEDLINE=96225458; PubMed=8640237;
 RA Miki Y., Katagiri T., Kasumi F., Yoshimoto T., Nakamura Y.;
 RT "Mutation analysis in the BRCA2 gene in primary breast cancers.";
 RL Nat. Genet. 13:245-247(1996).
 RN [9]
 RP VARIANT BC ASP-2089, AND VARIANT VAL-3412.
 RX MEDLINE=97294396; PubMed=9150152;
 RA Vetmanen P., Friedman L.S., Berola H., Sarantausta L., Pyrhoenen S.,
 RA Ponder B.A.J., Muhonen T., Nevanlinna H.;
 RT "A low proportion of BRCA2 mutations in Finnish breast cancer
 RT families.";
 RL Am. J. Hum. Genet. 60:1050-1058(1997).
 RN [10]
 RP VARIANT BC AND PANCREAS CANCER TRP-554.
 RX MEDLINE=98316775; PubMed=9654203;
 RA Ganguly T., Dhulipala R., Godmilow L., Ganguly A.;
 RT "High throughput fluorescence-based conformation-sensitive gel
 RT electrophoresis (F-CSE) identifies six unique BRCA2 mutations and an
 RT overall low incidence of BRCA2 mutations in high-risk BRCA1-negative
 RT breast cancer families.";
 RL Hum. Genet. 102:549-556(1998).
 RN [11]
 RP VARIANTS BC L-32; R-53; L-81; R-201; A-211; S-222 AND T-3118.
 RX MEDLINE=98272917; PubMed=9609997;
 RA Katagiri T., Kasumi F., Yoshimoto M., Nomizu T., Asaishi K., Abe R.,
 RA Tsuchiya A., Sugano M., Takai S., Yoneda M., Fukutomi T., Nanba K.,
 RA Makita M., Okazaki H., Hirata K., Okazaki M., Furutsuma Y.,
 RA Morishita Y., Iino Y., Karino T., Ayabe H., Hara S., Kajiwara T.,
 RA Houga S., Shimizu T., Toda M., Yamazaki Y., Uchida T., Kunitomo K.,
 RA Sonoo H., Kurebayashi J.-I., Shimotsuna K., Nakamura Y., Miki Y.;
 RT "High proportion of missense mutations of the BRCA1 and BRCA2 genes in
 RT Japanese breast cancer families.";
 RL J. Hum. Genet. 43:42-48(1998).
 RN [12]
 RP VARIANTS OC PRO-75; HIS-2502 AND HIS-3098.
 RX MEDLINE=20455732; PubMed=10486320;
 RA Gayther S.A., Russell P., Harrington P., Antoniou A.C., Easton D.F.,
 RA Ponder B.A.J.;
 RT "The contribution of germline BRCA1 and BRCA2 mutations to familial
 RT ovarian cancer: no evidence for other ovarian cancer-susceptibility
 RT genes.";
 RL Am. J. Hum. Genet. 65:1021-1029(1999).
 RN [13]
 RP VARIANTS HIS-289; ASN-372; ASP-991 AND VAL-3412.
 RX MEDLINE=99254821; PubMed=10323242;
 RA Li S.-L., Tseng H.-M., Yang T.-P., Liu C.-H., Teng S.-J.,
 RA Huang H.-W., Chen L.-M., Kao H.-W., Chen J.H., Tseng J.-N., Chen A.,
 RA Hou W.-F., Huang T.-J., Chang H.-T., Mok K.-T., Tsai J.-H.;
 RT "Molecular characterization of germline mutations in the BRCA1 and
 RT BRCA2 genes from breast cancer families in Taiwan.";
 RL Hum. Genet. 104:201-204(1999).
 RN [14]
 RP VARIANTS BC, AND VARIANTS.
 RX MEDLINE=99138688; PubMed=9971877;
 RA Wagner T.M.U., Hirtenlehner K., Shen P., Moeslinger R., Muhr D.,
 RA Fleischmann E., Concini H., Doeller W., Haid A., Lang A.H., Mayer P.,
 RA Petru E., Ropp E., Langbauer G., Kubista E., Scheiner O.,
 RA Underhill P., Mountain J., Stierer M., Zielinski C., Oefner P.;
 RT "Global sequence diversity of BRCA2: analysis of 71 breast cancer
 RT families and 95 control individuals of worldwide populations.";
 RL Hum. Mol. Genet. 8:413-423(1999).
 CC
 CC -!- FUNCTION: MAY PARTICIPATE IN A PATHWAY ASSOCIATED WITH THE
 CC ACTIVATION OF DOUBLE-STRAND BREAK REPAIR AND/OR HOMOLOGOUS
 CC RECOMBINATION.
 CC


```

CC CC -!- DEVELOPMENTAL STAGE: SPECIFIC TO LATE SPERMATOGENESIS.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL; Z22780; CAA80457.1; -
CC DR PIR; S35920; S35920.
CC DR PIR; B40713; B40713.
CC DR MIM; 603121; -
CC KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
CC FT NON_TER 1 1
CC FT DOMAIN 479 500 8 APPROXIMATE TANDEM REPEATS.
CC FT REPEAT 225 252 1.
CC FT REPEAT 253 289 2.
CC FT REPEAT 290 326 3.
CC FT REPEAT 327 364 4.
CC FT REPEAT 365 400 5.
CC FT REPEAT 401 438 6.
CC FT REPEAT 439 478 7.
CC FT REPEAT 479 500 8.
CC FT DOMAIN 548 598 PRO-RICH.
CC FT SEQUENCE 598 AA; 68034 MW; 062BAZE2D2AB61F7 CRC64;
CC CC
CC CC Query Match 4.5%; Score 136.5; DB 1; Length 598;
CC CC Best Local Similarity 23.3%; Pred. No. 0.14;
CC CC Matches 111; Conservative 72; Mismatches 205; Indels 89; Gaps 21;
CC CC
QY 2 GTTGLESLDGRGAAPTYSERLVPDPDLKEDVAMELERVGEDEEQMKRSEC 61
DB 70 GGPPLKDKSKKGGSTATNPSKQIVEKTRQNEAD-KTPLKSSHENESQSKSSSET 128
QY 62 NPLLQETIASAOGATAGTECKRSVPCGWERVVKORLFGKTAGRFDVYFI----SPOGLK 117
DB 129 NP-----ESQSKTVSKNSQ-----KDKKSKSKSKKTNTFELTKNPKKDL 171
QY 118 FRSKSLANYLKHNGTSLKPEDFTVLKRGKRSYKDCSMAALTSHLON--OSNNSN 175
DB 172 KRKSTNDPISICSENSL--NVDFLMVGQ-----SDDESINFDAWLRYSONNSKN 222
QY 176 WNLRTSKCKKDV--FMPPPSSSELQESR-GLSNFTSTHLLKKEDEGVDDV----- 223
DB 223 YSLKYTKYTKTKTKNAKSSDAESDSDAKDKKVKKKDKDKKDKKDKKDKKDKK 282
QY 224 -----NFRKVRKPKGVTKILKGIPIKTK-----GCRKSCSGFVQS 260
DB 283 ESGDSKDERKDTK-KDKKLLKDKKDKKDKKYPSTDTESGDADKARNDNRNLKASKND 341
QY 261 DSKRES--VCNKADAEPVAKQSOLDRTVCISDAGACETLSVTSENSLVKKKERLS 318
DB 342 DKKDKAKITFTDSESELESSESKQDKDKDK-----SKTDNKKK-VKNDSESD 392
QY 319 SGSNFCSEQKTSIITKFCSAKDSHNEKYEDT--FLESEIGTKVEVVERKEHLHTDIL 376
DB 393 ADSEPKGDSK-KGKDKKGGKDKKDKKDKKAKNAESTEMESDLELKKDKKHSKE---- 448
QY 377 KGSSEMDNCSPTKDFGCEKIFQEDTTPRQIERRKTSYFSSKYKNEALSPPRK 433
DB 449 KRGSKDKIK-KDKRKD-TESTDAEPDESSKTGF-KTSTKIKGSDTESSESLYKPKAG 502
CC CC
CC CC RESULT 9
CC CC BUD4_YEAST
CC CC ID BUD4_YEAST STANDARD; PRT; 1447 AA.
CC CC AC P47136;
CC CC DT 01-FEB-1996 (rel. 33, Created)
CC CC DT 01-OCT-1996 (rel. 34, Last sequence update)
CC CC DT 30-MAY-2000 (rel. 39, Last annotation update)

```

```

DE GN RUD SITE SELECTION PROTEIN BUD4.
OS BUD4 OR YJ052W OR J1905.
OC Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96295555; PubMed=8707826;
RA Sanders S.L., Herskowitz I.;
RT "The BUD4 prtein of yeast, required for axial budding, is localized
RL to the mother/BUD neck in a cell cycle-dependent manner.";
RN [2]
RP SEQUENCE FROM N.A.
RA Ramezani Rad M., Kirchrath L., Hollenberg C.P.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CO-ASSEMBLES WITH BUD3 AT BUD SITES. BUD4 AND BUD3 MAY
CC COOPERATE TO RECOGNIZE A SPATIAL LANDMARK (THE NECK FILAMENTS)
CC DURING MITOSIS AND THEY SUBSEQUENTLY BECOME A LANDMARK FOR
CC ESTABLISHING THE AXIAL BUDDING PATTERN IN G1.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN THE N-TERMINAL
CC SECTION DUE TO FRAMESHIFTS.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL; U41641; AAB17116.1; -
CC DR EMBL; Z49592; CAA89620.1; ALT_FRAME.
CC DR EMBL; Z49591; CAA89619.1; -
CC DR SGD; S0003852; BUD4
CC DR InterPro; IPR001849; PH.
CC DR Pfam; PF00169; PH; 1.
CC DR SMART; SM00233; PH; 1.
CC DR PROSITE; PS50003; PH_DOMAIN; 1.
CC KW ATP-binding.
CC FT DOMAIN 181 189 ASP/GLU-RICH (ACIDIC).
CC FT NP_BIND 1175 1182 ATP (POTENTIAL).
CC FT DOMAIN 1302 1413 PH.
CC FT CONFLICT 291 328
CC FT CONFLICT 340 340
CC FT SEQUENCE 1447 AA; 164485 MW; 1D056B9FF1B7067B CRC64;
CC CC
CC CC Query Match 4.4%; Score 133; DB 1; Length 1447;
CC CC Best Local Similarity 19.5%; Pred. No. 0.73;
CC CC Matches 135; Conservative 105; Mismatches 269; Indels 184; Gaps 30;
CC CC
QY 1 MGTGLESLSLG-----DRGAAPTYSERL-----VDPDPDLKEDVAM 41
DB 74 MSTISNESLGLLRNVNSELSESPAAVHOERTKNSVANGALGHANSPKVLNLLK--NMAQ 131
QY 42 ELERVGEDEEQMKIKRSS-----ECNPLIQETIASAOGATAGTECKRSVPCGWERVV 94
DB 132 DIDKLARDEKPKVLSLSSPLKFTLKSTQPLLSYPESPIH-----RSSIE-----I 176
QY 95 KQRLFKTAGRFDVYFI---SPOGLKRSKSLANYLKHNG---ETSLKPEDFTVLKSK 148
DB 177 ETNYDDEDEEDDAYTCLTQSTQILHSPSRIPITNAVSINKLNLDFLTPNPNESDKSLVD 236
QY 149 RGIKSKYKD-----CSMAA-----LTSHLONNSNNLNLRTRSKCKK 186
DB 237 TSVDSGTGRELDTKTIPELPCKNSSTFMPVDEKCNLPKLLNTSNNSHDSRSPTASVE 296
QY 187 DV----FMPSPSSSELSQESKGLSN--FTSHLLKED-----EGVDYVNRKVRKP 231

```

```
Db 297 DLNISTNLPAGDSQ-----NNPVTDDADALITENDVVRDLQNMHEHIDDAFDEKVKLD 349
QY 232 KG-----KVTIL---KGIPKTKGCRKSCSGFVQOSKRES-----VNKA 271
Db 350 EGCNSNEPVTFLGENDTRISIVYSGKGTANVQEFQSDSLAHSEPKFKDLNATSDVWNE- 408
QY 272 DAESPVAQKSQLDRTVCISDAGAGETLSVTSEENSLVKKKRSLSGSGNFCSEOKTSG 331
Db 409 DKETDANISTSKSESYIAD-----YKVTROEDWDTK-----LHQSEHANEQ--PA 455
QY 332 TINKFCSAKSEHNEKYEDTFLSEIEIGTKVVEVVERKEHLTDILK-----RGSEM-- 382
Db 456 IIPQDSSEETFTLNNESEFORNKDGEYRIVQHEESLYGORTKSPENLINGSEIGV 515
QY 383 -----DNCSPTKDFTEKIFQEDTTPRTQIERKTSL----- 416
Db 516 DHGEAAVNEPLAKTSABEHLSDSCEDQSVSEARNKDRIEKEVETKDNETIEKDESE 575
QY 417 YFSSKYNKEALSPRRKAFKWTPTPRSPFNVLQVETLPHDPW-----KLLIATIFLNRTSGK 472
Db 576 YHKVEENEPEHVPLPLPW-----EELQNEPFDIDENDTNSIDITRSKWP 625
QY 473 MAIPVLWKFLE---KYPSAEVARTADMRDVSSELLKPLGLYDLR--AKTIKVFSEYILTKQ 527
Db 626 SDYISNIHQEEIKSNPESIANQSFSQSSITTASTVDSKDKNGSTSPFKPKRIVRSR 685
QY 528 WKY---PIELHGLGKYGNDYRIFCVNEWKVQHP 558
Db 686 RIYNPKRSVSLNYDNEY-ILSNSEWNALDP 717

RESULT 10
BIRL_SCHPO
ID BIRL_SCHPO STANDARD; PRT; 997 AA.
AC 014064;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE BIRL PROTEIN (CHROMOSOME SEGREGATION PROTEIN CUT17).
GN BIRL OR CUT17 OR SPCC962.02C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID:4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Morishita J., Matsusaka T., Yanagida M.
RT "Fission yeast cut17 is required for chromosome segregation.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RT Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=99398681; PubMed=10468581;
RA Uren A.G., Beilharz T., O'Connell M.J., Bugg S.J., van Driel R.,
RA Vaux D.L., Lithgow T.;
RT "Role for yeast inhibitor of apoptosis (IAP)-like proteins in cell
RT division";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10170-10175(1999).
CC -!- FUNCTION: SEEMS TO ACT IN THE PLEIOTROPIC CONTROL OF CELL
CC DIVISION. MAY PARTICIPATE IN CHROMOSOME SEGREGATION EVENTS.
CC -!- SIMILARITY: CONTAINS 2 BIR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AR031034 : BAA83415.1; -
DR EMBL: AL031323 : CAA20434.1; -
DR InterPro: IPR001370 : BIR.
DR Pfam: PF00653 : BIR; 2.
DR SMART: SM00238 : BIR; 2.
DR PROSITE: PS01282 : BIR_REPEAT_1; FALSE_NEG.
DR PROSITE: PS0143 : BIR_REPEAT_2; 2.
KW Cell division; Repeat.
FT REPEAT 25 99 BIR 1.
FT REPEAT 120 194 BIR 2.
FT DOMAIN 80 83 POLY-ASP.
FT DOMAIN 312 319 POLY-ASP.
FT DOMAIN 487 490 POLY-SER.
SQ SEQUENCE 997 AA; 112579 MW; 952A6BAFA5C489F4 CRC64;

Query Match 4.3%; Score 131.5; DB 1; Length 997;
Best Local Similarity 20.1%; Pred. No. 0.56;
Matches 129; Conservative 119; Mismatches 241; Indels 153; Gaps 33;

QY 20 VTSSERLVPDPNDLRKEDVAMEL---ERVGEDEEQMMIKRS---SECNPLLQEPISA 72
Db 416 ISSSVSVGKEQNHTKQ-VAIETPEQKVEKEDHLNLQSGSFIESTKQPISSKPTSS 474
QY 73 QFCATAGTECRKSPVCGWVRVVKQRLFGKTAGREDYVIFISPGQLKPKRSKSSLIANYLHKG 132
Db 475 PDMTDAATGGRVSSSRFRDIQTNPFRST-----IDFSNISKKNSEAN--DEND 526
QY 133 ETSLK---PE---DEDFTVLSKRGIKSRKYDC-----SMAAL-----TSHLQNSNN 173
Db 527 ETNLKIPIDPEKRRKFEVLQSKNLVSSSTEDSHPVKVETDSQTAHVSKFDELENKSM 586
QY 174 SNWNLRTSRCKKD---VEMPPSSSSSELOES-----RGLSNFTSTHLLKDEGVDD 222
Db 587 SEQSLQSSSENDKPLDILPLAIKRDNLVSGVLEKKGKSTSK---TKFDTSIDV 643
QY 223 VNFKVRKPKGVITLKGIPDKTKKGRKSCSGFVQ-----SDSKRESV 267
Db 644 F----IEKPKTEIS--EVLPEEKRAKICDESQTVRVSIDRGVTKTROVSSPVSDKSEN 697
QY 268 CNKADAES-----EPVAKSOLDRTVCISDA--GACGETLSVTSEENSLVKK 312
Db 698 -NHEEANSNGHTVMNVHSSLDPPQIVQPNLESGSYLKDIPDRWVGNSKVTFOEDDINS 756
QY 313 KERSLSSGNSFCSEQKTSIGIKFCSAKDEHNEKYEDTFLSEIEIGTKVVEVVERKEHLH 372
Db 757 KLSKNQTVAEAVNTETSDKLOKEANHELENIEKIEEKLTVEYDKVSLSDAFPQDEIKNS 816
QY 373 TDILKRGSEMDNNSPTRKDFTEKIFQEDTTPRTQIERR-----KTSLYFSSKY-NKEAL 427
Db 817 RTSVQNGTRSVSKNTPEKE---TKVDKIDNVSKKDVETSPGSCETSSAFKTYAEKVT 872
QY 428 S----PPRRKAFKK--WTPPRSPLNLVQETLPHDPWKLILLATIFLNKTSKMAIPVLWKL 482
Db 873 SINLPSVRKPLDESYDHSISPFDPCLQSSFLAP-----QTPVKSCKHALPLV--- 919
QY 483 EKYPSEAEVARTADWR--DVSELLK-----PLGLYDLRAKTIVKFSDEYLTQW-KYPIEL 534
Db 920 -----EANAPWPEIDFSSLLSESPVPNPVFNPKLSEKEL-----DMTVEQWIKF---- 963
QY 535 HGIGKYGNDYSYIFCVNEWKVQHPDHKLNYKHYHDWLWENHEK 576
Db 964 ---MVAK-----CAKEFEACEE--KI-----EWLEECKR 989

RESULT 11
ANK2_HUMAN
ID ANK2_HUMAN STANDARD; PRT; 3924 AA.
AC Q01484; Q01485;
DT 01-APR-1993 (Rel. 25, Created)
```


Db 9 VTPSGKLVSETVKNEKTSDETLK-----AKGKEIKVNSRREDFKQKQP 59
 QY 165 SH-----LQNSNNSMNLRTRCKCKDVMPSPSS-----SELQE----- 200
 Db 60 SKKKRIIYDSSESETLOVKNNAKPPKPLPVSSPKGLSKRPDPVYISETDEEDFEMCK 119
 QY 201 -----SRGLSNFTSTHL-----LLKDEGVDDVNFVKRPKPKGVILKGIPTKTKGCR 251
 Db 120 KAASKSKNGRSTNSHLGTSNKKKEE-----NTRTKNKPJS-----PIKLPTSVL 166
 QY 252 KSC-SGFVQSD-----SKREVCNKAD-----AESPEVAOKSOLDRTV-----CISDAGAG 297
 Db 167 DVFGTGVSORSNKKMVASKRKLSQNTDSGLNDEAIAKQLDDEDAELERHLEDEEFA 226
 QY 298 ETLVSTSENSLVKKERSLSSGNSFCSEQ-----KTSGLI----- 333
 Db 227 RTLAMDDEPK-TTKARKDTEAGETFSVQANLSKAERHYPHYKVTQAQVDSRKSYSR 285
 QY 334 --NKFCASDSE-EHNEKYEDTF--LESEIGTKVFEVERKEH-LHTDILKRGSEMDNNC- 386
 Db 286 KQSKYESSKESQOHSKSSADKIGEVSSPKASSKLAIMKRKESSYKEIEPVASKRKENAI 345
 QY 387 -----SPTKDFTEKIFQEDTTPRTQIERKTSLY-FSKYNKEALSPPR 431
 Db 346 KLGKETKTPKTKSSPAKK-----ESVSPDS-----EKKRTNYQARYSLNREG----- 390
 QY 432 RKAFKWTTPRSPFNLVQ-----ETLPHDPWKLIIAT-----IFL 466
 Db 391 PKALGSKEIPKGAENCLEGLIFVITCVLESIERDEAKSLIERGGKVTGNSKKTNYLVLM 450
 QY 467 NRTSGK-----MAIPVLWKELE-----KYPSEAVARTADWRDVSLE----- 502
 Db 451 GRDSGQSKDKAAALGTIKIDEDGLNLIRTPMGKKSKYEIAVETEMKESKERTPOKN 510
 QY 503 -----LKPGLYDLRAKTIIVKPSDEVLTKQWK----- 529
 Db 511 VQGRKISPSKESKSRPTSKRSLAKTIKETDVF-----WKSDFKEQVAETSD 566
 QY 530 -----YPIELHG-IGKYGNDSDYRIFVNE-----WKQVH 557
 Db 567 SKARNLADDSSENKVENLLWVDKYKPTSLKTIIGQOGDQS-----CANKILLMLRNQKWS 622
 QY 558 PEDHKLNY 566
 Db 623 SEDKHKSKF 631

RESULT 13
 YF16_YEAST
 ID YF16_YEAST STANDARD; PRT; 1233 AA.
 AC F43597;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE HYPOTHETICAL 137.7 KDA PROTEIN IN UGSI-FAB1 INTERGENIC REGION.
 GN YF016C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=95400292; PubMed=7670463;
 RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
 RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
 RA Yamazaki M., Tashiro H., Eki T.;
 RT "Analysis of the nucleotide sequence of chromosome VI from
 RT Saccharomyces cerevisiae."
 RL Nat. Genet. 10:261-268(1995).
 CC -i- SIMILARITY: SOME, TO MAMMALIAN NEUROFILAMENT TRIPLET M PROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: D50617; BAA09255.1; -
 DR SGD: S0001912; YF016C.
 KW Hypothetical protein.
 SQ SEQUENCE 1233 AA; 137697 MW; C8A7CD2C6F0892F6 CRC64;

Query Match 3.9%; Score 120.5; DB 1; Length 1233;
 Best Local Similarity 18.9%; Pred. No. 3.7;
 Matches 121; Conservative 91; Mismatches 228; Indels 199; Gaps 25;

QY 7 ESLSLGRGAAPTVTS-----SERLVDPDPND-----LRKEDVAMELERYCEDEEQM 53
 Db 546 DNLVLEDEAEAPTQENKPTQEVGVIDIPDAPRDDVLEAVEKNIIPEDLEVAKEDEGE 605
 QY 54 MIKRSSECNPLQSEPI-----ASAQFGATA- 78
 Db 606 QVK-----LDEPVAMKDDKIAMRGASISSEDMKKKQEGTAELSNEKAKKEVDETAR 657
 QY 79 ---GTECRKSVPCGWERVYKQRLFGKTAGRDYVYFISPOGLKPRS----- 120
 Db 658 ESAGGEVEKSKTPSPKVKR-----CTSGR-----PEDLQINERDPEVLKEDVRVPD 706
 QY 121 ---KSSLVNLYHKNE-----TSL-----KPEDFDFTV 145
 Db 707 EDVPEIATTIENSEEEDPKSQVQISTEQAEITQKMDGVGVTTFSEKEEEKPKRFEIT- 765
 QY 146 LSKRGIKSRKDCSMAALTSHLQNSNN-SNNWNLRTSKCKKDVFMPPSSSELOESRGL 204
 Db 766 --QEGDKITCKD-----TNHEGAEATEAASSENSKASDVGTAEKYIDPSSSVKKT-- 814
 QY 205 SNFTSTHLLKDEGVDDVNFVKRPKPKGVITLKGPIKTKKCKKSCGFVSDSKR 264
 Db 815 -----EEDAEVENSEKTEFIKVAE-----LENDAPKAE-----VTAELENK 852
 QY 265 ESVCKNAIDAESEPAOKSOLDRTVCIS-----DAGAGETLSVTSSENSLVKKKRSLS 318
 Db 853 ENEDVEVTEDEAEVENSEKTEFIKVAELNLDPAKAEVTAELNKENEDVEVAATSKE 912
 QY 319 SGSNFCSEQKTSGLINKFC-----SAKDSHENEKYEDTFLSEBEIGTKVEVVERKEHLH- 372
 Db 913 DIETKCEPAETPIEDGTCTEAEVSKKDAEAVTK-EDENMENSKIAPALKDVTGDQDIDD 971
 QY 373 --TDILKRGSEMDNNSPTKDFTG-----EKIFOEDTIPRTOIERKTSLYFSSKY 422
 Db 972 INISDEFORTVELPELEKQDKIKNKGDEKLEVEETKETSPLDVLVEENITEKNEIKQ 1031
 QY 423 NKEALSP---PRKAFKKWTTPRSPFNLVQETFLFHPDWKLLIATIF---LNRTSGKMAIP 476
 Db 1032 EEEVSQLDNFNETESISKEAPNNDENFEGDQSTRENPKASADDIPKIDILDET- 1085
 QY 477 VLWKELEYPSAEVARTADWRDVSLELLKPLGLYDLRAKT 515
 Db 1086 ---EFLEQ-----LKIVDDSENLALQSLDAKDDSTTQT 1115

RESULT 14
 AC15_MOUSE
 ID AC15_MOUSE STANDARD; PRT; 1131 AA.
 AC P35601;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ACTIVATOR 1 140 KDA SUBUNIT (REPLICATION FACTOR C LARGE SUBUNIT) (A1
 DE 140 KDA SUBUNIT) (RF-C 140 KDA SUBUNIT) (ACTIVATOR 1 LARGE SUBUNIT)
 DE (A1-PI45) (DIFFERENTIATION SPECIFIC ELEMENT BINDING PROTEIN)

DE (ISRE-BINDING PROTEIN).
GN RFCL OR RECC1 OR IBF-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BA1B/C;
RX MEDLINE=94089669; PubMed=8265586;
RA Burdello P.D., Utani A., Pan Z., Yamada Y.;
RT "Cloning of the large subunit of activator 1 (replication factor C)
RT reveals homology with bacterial DNA ligases.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:11543-11547(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94158835; PubMed=8114700;
RA Luckow B., Bunz F., Stillman B., Lichter P., Schuetz G.;
RT "Cloning, expression, and chromosomal localization of the
RT 140-kilodalton subunit of replication factor C from mice and
RT humans.";
RL Mol. Cell. Biol. 14:1636-1634(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS;
RX MEDLINE=95388065; PubMed=7659092;
RA McGehee Habener J.F.;
RT "Differentiation-specific element binding protein (DSEB) binds to a
RT defined element in the promoter of the angiotensinogen gene required
RT for the irreversible induction of gene expression during
RT differentiation of 3T3-L1 adipoblasts to adipocytes.";
RL Mol. Endocrinol. 9:487-501(1995).
RN [4]
RP SEQUENCE FROM N.A.
RA Haque S.J.;
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-565 FROM N.A.
RC STRAIN=LAF1;
RA Lossie A.C., Haugen B.H., Wood W.M., Camper S.A., Gordon D.F.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 354-528 FROM N.A.
RX MEDLINE=90229765; PubMed=1691767;
RA Haque S.J., Kumar A., Fischer T., Rutherford M.N., Williams B.R.;
RT "Evaluation of inter- and intramolecular primary structure homologies
RT of interferons by a Monte Carlo method.";
RL J. Interferon Res. 10:31-31(1990).
CC -!- FUNCTION: THE ELONGATION OF PRIMED DNA TEMPLATES BY DNA POLYMERASE
CC DELTA AND EPSILON REQUIRES THE ACTION OF THE ACCESSORY PROTEINS
CC PCNA AND ACTIVATOR 1. THE 140 SUBUNIT BINDS TO THE PRIMER-
CC TEMPLATE JUNCTION.
CC -!- SUBUNIT: HETEROPENTAMER OF SUBUNITS OF 140/145, 40, 38, 37, AND
CC 36.5 KDA THAT FORMS A COMPLEX WITH PCNA IN THE PRESENCE OF ATP.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE ACTIVATOR 1 140 KDA SUBUNIT FAMILY.
CC -!- SIMILARITY: CONTAINS 1 BCT DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U01222; AAA21643.1;
DR EMBL: X72711; CAA51260.1;
DR EMBL: U36441; AAA79698.1;
DR EMBL: U07157; AAC52140.1;
DR EMBL: U15037; AAB60452.1;
DR MGD; MGI:97891; Reccl.

DR InterPro: IPR001939; AAA_subfam.
DR InterPro: IPR001357; BRCT.
DR InterPro: IPR000862; RFC.
DR Pfam: PF00004; AAA; 1.
DR Pfam: PF00533; BRCT; 1.
DR SMART: SM00292; BRCT; 1.
DR PROSITE: PS0172; BRCT; 1.
KW DNA replication; ATP-binding; Transcription regulation; DNA-binding;
KW Activator; Nuclear protein.
FT DOMAIN 399 477 BRCT.
FT NP_BIND 635 642
FT ZN_FING 734 751
FT DOMAIN 1104 1108
FT DOMAIN 354 528
FT CONFLICT 66 66
FT CONFLICT 187 187
FT CONFLICT 254 254
FT CONFLICT 559 559
FT CONFLICT 614 614
FT CONFLICT 945 945
FT CONFLICT 1071 1071
FT CONFLICT 1104 1104
FT CONFLICT 1131 1131
SQ SEQUENCE 1131 AA; 125984 MW; A6F4F970A7F9EE94 CRC64;
Query Match 3.9%; Score 120; DB 1; Length 1131;
Best local Similarity 20.1%; Pred. No. 3.6;
Matches 120; Conservative 76; Mismatches 182; Indels 218; Gaps 30;
QY 119 RSKSSLLANYLHNKGETSLK-----PEDFDFTVLSKRGKISRYKDCSMAALTSHLQN 169
Db 82 KKSEKLSLSYKPGKVSQKDPVTVSETDEDDF-VCKKAASKSKENGVS-----TNSYLG 136
QY 170 QSN--NSNNLRTSRCKCKDVFMPSS-----SSSELOESR 202
Db 137 TSNVKNNEENVKTKNKLPTSTVLDVFGTESVQSGKKMVTSKRKESQNTEDSR 196
QY 203 GLSNFTSTHLLKKEGV-----DDVNFVKV-----RKPQKVTILKIPKTKKCRK 252
Db 197 LNDEAIAKQQLDDEDAELERQLHIEDEFARTLALLDEPK-----IKARKDSEE 246
QY 253 SCSGF--VQSD---SKRESVCNKAD-----AESEPVQV--KSLQDRTV 288
Db 247 GEESFSSVQDDLSKAKQKSPNKAELFSTARKTYSAPAKHGKGRASEDAKOPCKSAHRKEA 306
QY 289 CISDAGACGETLSVTSENS-----LVKKERSLSSGNSFCSEQKT--SGIINKFCSA 339
Db 307 CSSPKASAKLALMAKAESESYNETELLAARRKESATEPKGKTKTPKTKVSPTKRESVSP 366
QY 340 KDSEH---NEKYDTFLESE---EIGTK-----VEVVERKEHLHT 373
Db 367 EDSKKRTNYQARYSYLNREGPKALGSKTEPKGAENCLGLTFTVITGVLESIERDE-AKS 425
QY 374 DILKRGSEMDNCS-----PTRKD---FTGEKIFQD-----TIP----- 405
Db 426 LIERYGKVTQNSVSKTNYLVMGDRSGQSKDAAALGTLKILDEGLDILRTMPGKRK 485
QY 406 -----RTQIERRTKSLTFSSKYN---KEALSPRR-----RAFKKWTPPRSPFNVLQETLF 453
Db 486 YEMAAEAEMKKEKLETPKNDQGRKISPAKKESEKSKKLTLLKNSPMKAVKKEAS 545
QY 454 HDPKLLIATIFLRTSGKMAIPVLWKFLEKYPSAEVARTADWRDVSSELLKPLGLYLRA 513
Db 546 TCPRLGVDKETHGRNKNKEEC-LLW--VDKY-----KPAKLKNI-- 582
QY 514 KTVKFSDEYLTQWKYPILHGLGICKYGNDSYRIFCVNE-----WKQVHPDHHK 562
Db 583 -----IGQQDQS-----CANKLLRLNRLNWKSSPEEKK 611
RESULT 15
YM67_YEAST

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 14:06:25 ; Search time 134.53 Seconds
(without alignments)
344.669 Million cell updates/sec

Title: US-09-724-296-37

Perfect score: 1662

Sequence: 1 MIFRFGFVSNAMSLWDSPA.....ELSSIRGVKRGIGGALQWKS 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1001	60.2	317	2 Q9L4F0	Q9L4F0 bacillus ce
2	1001	60.2	325	2 Q9L4E9	Q9L4E9 bacillus ce
3	484.5	29.2	322	2 Q9K9P8	Q9K9P8 bacillus ha
4	359.5	21.6	599	3 Q10988	Q10988 schizosacch
5	345	20.8	656	3 Q01408	Q01408 neurospora
6	293.5	17.7	305	2 Q9S0M9	Q9S0M9 deinococcus
7	293.5	17.7	326	2 Q9RTE6	Q9RTE6 deinococcus
8	109.5	6.6	695	3 Q9HGM8	Q9HGM8 schizosacch
9	104.5	6.3	315	2 Q9X2Y9	Q9X2Y9 bacillus an
10	104.5	6.3	1690	3 Q9P977	Q9P977 candida alb
11	104	6.3	2272	5 Q9VZS7	Q9VZS7 drosophila
12	101.5	6.1	1690	3 Q9P411	Q9P411 candida alb
13	101.5	6.1	1690	3 Q9HFT8	Q9HFT8 candida alb
14	101	283	1	P95930	P95930 sulfolobus
15	100	6.0	823	2 Q9KB81	Q9KB81 bacillus ha
16	98	5.9	331	2 Q4S152	Q4S152 bacteroides
17	98	5.9	505	5 Q9NGC2	Q9NGC2 triboilium c
18	97.5	5.9	1744	11 Q9R095	Q9R095 rattus norv
19	96.5	5.8	969	3 Q14014	Q14014 schizosacch

20	96.5	5.8	1760	5 Q9VLT3	Q9VLT3 drosophila
21	95	5.7	775	5 Q45519	Q45519 caenorhabdi
22	94.5	5.7	553	5 Q19702	Q19702 caenorhabdi
23	94.5	5.7	1145	10 Q04146	Q04146 acetabulari
24	94	5.7	553	2 Q9FCX5	Q9FCX5 clostridium
25	93.5	5.6	805	5 Q17235	Q17235 caenorhabdi
26	93.5	5.6	1066	5 Q9TZX4	Q9TZX4 caenorhabdi
27	93	5.6	674	5 Q20833	Q20833 caenorhabdi
28	92.5	5.6	234	4 Q9Y2M6	Q9Y2M6 homo sapien
29	92.5	5.6	601	4 Q75226	Q75226 homo sapien
30	92.5	5.6	674	4 Q9UBU7	Q9UBU7 homo sapien
31	92	5.5	261	12 Q9YFX9	Q9YFX9 maize strea
32	92	5.5	428	1 Q9UYK0	Q9UYK0 pyrococcus
33	92	5.5	921	2 Q9LZ27	Q9LZ27 heisseria m
34	91.5	5.5	237	2 Q9PCN0	Q9PCN0 xyliella ras
35	91.5	5.5	805	5 Q93719	Q93719 caenorhabdi
36	91.5	5.5	1035	4 Q9H2F7	Q9H2F7 homo sapien
37	91.5	5.5	1052	11 Q9ERA1	Q9ERA1 microtus ar
38	91.5	5.5	1409	5 Q08461	Q08461 drosophila
39	91	5.5	558	10 Q9STH1	Q9STH1 arabidopsis
40	91	5.5	852	2 Q9K940	Q9K940 bacillus ha
41	91	5.5	1020	10 Q9ZVT9	Q9ZVT9 arabidopsis
42	90.5	5.4	279	5 Q17470	Q17470 caenorhabdi
43	90.5	5.4	448	5 Q18125	Q18125 caenorhabdi
44	90.5	5.4	553	5 Q19881	Q19881 caenorhabdi
45	90.5	5.4	625	5 Q44733	Q44733 caenorhabdi

ALIGNMENTS

RESULT 1					
Q9L4F0					
ID Q9L4F0	PRELIMINARY;	PRT:	317 AA.		
AC Q9L4F0;					
DT 01-OCT-2000 (Tremblrel. 15, Created)					
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)					
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)					
DE PUTATIVE UV-ENDONUCLEASE.					
GN YWJD.					
OS Bacillus cereus.					
OC Bacteria; Firmicutes; Bacillus/Clostridium group;					
OC Bacillus/Staphylococcus group; Bacillus.					
OX NCBI_taxID=1396;					
RN [1]					
RP SEQUENCE FROM N.A.					
RC STRAIN-ATCC 14579 TYPE STRAIN;					
RX MEDLINP=20055637; PubMed=10589720;					
RA Okstad O., Gominet M., Purnelle B., Rose M., Lereclus D., Kolsto A.B.;					
RT "Sequence analysis of three Bacillus cereus loci under PlcR-regulated					
RT genes encoding degradative enzymes and enterotoxin.";					
RL Microbiology 145:3129-3138(1999).					
DR EMBL; AJ243712; CAB69813.1;					
KW Endonuclease.					
SQ SEQUENCE 317 AA; 36990 MW; A6AC35F5800A22C3 CRC64;					

Query Match	60.2%;	Score 1001;	DB 2;	Length 317;	
Best Local Similarity	58.2%;	Pred. No. 2.7e-74;			
Matches 185;	Conservative 55;	Mismatches 74;	Indels 4;	Gaps 2;	
OY 1	MIFRFGFVSNAMSLWDSPAKTLTFARYSKLSKTERKALLTVTKANLRNTRMRTLHYIIG	60			
Db 1	MLIRGGYSHAMALWDCSPAKTMTFTSPKLSKOEKDKLYHVRQNLERTIRILYNIA	60			
OY 61	HGIPLYRESSSTIVPLATHPDVMDVFTPFQKEFRIGELVKTHOLRTSFHPNQFTLTSP	120			
Db 61	HEIPLYRLSSSTIVPLATHPEVEFDYIGVFTPLWRKIGALIKEHNLRISFHPNQFTLTSD	120			
OY 121	KESVTKNAVOTMAYHYRMLEANGIADRSVINIHGGVGNKDTATQAFQNIKOLPOEIK	180			
Db 121	KPHITTNALTMDTYHYKILDAIGTADSSYINIHVGAGYGNKEKATERPHENIKKLPHTHIK	180			

```

QY 181 ERMTELENDKTYTTEETLQVCEQEDVPVDFHHFYANPDHDLNVALPRMIKTWERIG 240
Db 181 KOMTLENDKTYTTEETLQVCEQEDVPVDFHHFYANPDHDLNVALPRMIKTWERIG 239
QY 241 LOPKVHLSPPKSEQAIRSHADYVDANFL---LERFROWGNTNIDPMIBAKQKALLRLMD 297
Db 240 ISPKVHSSPSEKFEFRAHAEYIDLEFKPLHTAKKHHNFDMIESKQKDLALFOLID 299
QY 298 ELSSIRGVKRIKGGALOW 315
Db 300 ELSAIRGIRKISGAMLOW 317

RESULT 2
Q9L4E9 PRELIMINARY: PRT; 325 AA.
AC Q9L4E9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PUTATIVE UV-ENDONUCLEASE.
GN YWJD.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14579 TYPE STRAIN;
RX MEDLINE=20053637; PubMed=10589720;
RA Oktad O., Gominet M., Purnelle B., Rose M., Lereclus D., Kolsto A.B.;
RT "Sequence analysis of three Bacillus cereus loci under P1cr-regulated
RT genes encoding degradative enzymes and enterotoxin.";
RL Microbiology 145:3129-3138(1999).
DR EMBL: AJ243712; CAB69814.1; -.
KW Endonuclease.
SQ SEQUENCE 325 AA; 37900 MW; 07FBDD1FA580653 CRC64;

Query Match 60.2%; Score 1001; DB 2; Length 325;
Best Local Similarity 58.2%; Pred. No. 2.8e-74;
Matches 185; Conservative 55; Mismatches 74; Indels 4; Gaps 2;

QY 1 MIFRFGFVSNAWLDASPAKTLTFARYSKLSKTEREALLTVTKANLRNTMRTLHIIG 60
Db 9 MLIRFGYVSHAMALWDCSPAKTMTFTSFKLSKQEREDKLVHYRQNLHTIRTLHYNIA 68
QY 61 HGIPYRFSSTIVPLATHPDVMDVFTVPQKEFREIGELVKTHOLTSFHPNQFTLFTSP 120
Db 69 HEIPYRLSSSIVPLATHPEVEFDYIGVFTPLWRKIGALIKEHNLRSFHPNQFTLFTSD 128
QY 121 KESVTKNAVTDMAHYRMLEANGIADRSVINIHGGAYGNKDTATATQPHONIKOLPOEIK 180
Db 129 KPHITNAVTDMAHYRMLEANGIADRSVINIHGGAYGNKDTATATQPHONIKOLPOEIK 188
QY 181 ERMTELENDKTYTTEETLQVCEQEDVPVDFHHFYANPDHDLNVALPRMIKTWERIG 240
Db 189 KOMTLENDKTYTTEETLQVCEQEDVPVDFHHFYANPDHDLNVALPRMIKTWERIG 247
QY 241 LOPKVHLSPPKSEQAIRSHADYVDANFL---LERFROWGNTNIDPMIBAKQKALLRLMD 297
Db 248 ISPKVHSSPSEKFEFRAHAEYIDLEFKPLHTAKKHHNFDMIESKQKDLALFOLID 307
QY 298 ELSSIRGVKRIKGGALOW 315
Db 308 ELSAIRGIRKISGAMLOW 325

RESULT 3
Q9K9P8 PRELIMINARY: PRT; 322 AA.
AC Q9K9P8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)

```

```

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE BH2597 PROTEIN.
GN BH2597.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kubara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001516; BAB06316.1; -.
KW Complete proteome.
SQ SEQUENCE 322 AA; 36900 MW; 4B0DC9291CD1E7B4 CRC64;

Query Match 29.2%; Score 484.5; DB 2; Length 322;
Best Local Similarity 32.2%; Pred. No. 6.7e-32;
Matches 106; Conservative 67; Mismatches 133; Indels 23; Gaps 6;

QY 1 MIFRFGFVSNAWLDASPAKTLTFARYSKLSKTE---RKEALLTVTKANLRNTMRTLHI 57
Db 1 MRIOGFIYVAMSELANASPKTMTATQFEKIEDHAGLRK--LERIAKTNLHNCILRLKH 58
QY 58 IIGHGIPYRFSSTIVPLATHPDVMDVFTVPQKEFREIGELVKTHOLTSFHPNQFTL 116
Db 59 NLAYQISFSEFLSKLPLVNLHPLTEGKYELALAEALQAVGEFASEHOMRIDFHPDFV 118
QY 117 FTSPKESVTKNVTDMAHYRMLEANGIADRSVINIHGGAYGNKDTATATQPHONIKOLP 176
Db 119 LNSEAKEITRRSLQTLHYKLLKGMEDIPRRCVLHVGGKKKGVEAGLEQFIENTASIP 178
QY 177 QETKERMTLENDKTYTTEETLQVCEQEDVPVDFHHFYANPDH---ADLNVALP 230
Db 179 KSLLSMIMLENDKSYTIDDLVYLGEKLAIPVVDLHH-----HDVLHRSKSLQETWQ 231
QY 231 RMKTWERICLOPKVHLSPPKSEQAIRSHADYVDANFLERFROWGT---NIDPMIEAK 286
Db 232 RIVATWEDSPKPKHLSPLSGEDDPRHHDYINADRFATLHEICADAVDHLHVMIEAK 291
QY 287 QKDKALLRLMDELSSIRGVKRIKGGALOW 315
Db 292 KKDALLFQLMKDLAEYDEITVVSXSAVEF 320

RESULT 4
Q10988 PRELIMINARY: PRT; 599 AA.
AC Q10988; p87339;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE UV-ENONUCLEASE.
GN UVDE OR UVE1+.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sp972;
RX MEDLINE=9618860; PubMed=8614629;
RA Takao M., Yonemasa R., Yamamoto K., Yasui A.;
RT "Characterization of a uv endonuclease gene from the fission yeast
RT Schizosaccharomycetes pombe and its bacterial homolog.";

```

Nucleic Acids Res. 24:1267-1271(1996).

RP SEQUENCE FROM N.A.
 RX MEDLINE=97175806; PubMed=9023111;
 RA Davey S., Nass M.L., Ferrer J.V., Sidik K., Eisenberger A.,
 RA Mitchell D.L., Freyer G.A.;
 RT "The fission yeast UVDR DNA repair pathway is inducible.";
 RN Nucleic Acids Res. 25:1002-1008(1997).
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA Lyne M., Rajadream M.A., Barrell B.G., Lucas M., Gaillardin C.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; D78571; BAAL1415.1;
 DR EMBL; U78487; AAC49664.1;
 DR EMBL; AL023859; CAA19577.1;
 KW Endonuclease.
 SQ SEQUENCE 599 AA; 68816 MW; C09ED7219D5895BE CRC64;

Query Match 21.6%; Score 359.5; DB 3; Length 599;
 Best Local Similarity 29.7%; Pred. No. 2.8e-21;
 Matches 83; Conservative 60; Mismatches 111; Indels 25; Gaps 7;

QY 34 TERKEALLVTKANLNTW---RTLHYLIIGHGIPLYRFSSSIVPLATHPDVMMDFVTFP- 89
 DB 276 TOROGLESVKQGTQNVLDLIKLEWNNFNGIHPMRVSSDLFPFASH--AKYGYLLEFA 333
 QY 90 KBREFEIGELVTHQURTSFHPNQFTLTSPKESVTKNVDMAYHYRMLEAMGIADR-- 147
 DB 334 QSHLEEVGLKANKYHRLTMHPQVQTOIASPREVVVDSAIRDLAYHDEILSRMKLNEQLN 393
 QY 148 --SVINIHGGAGKNDTATQAFHONIKOLPOEIKERMTLENDDKTYTTEETLQVCEQED 205
 DB 394 KDAVLIHLGGTPEGRKETIDRFKRYQRLSDSVKARLVLENDVSVQDILLPLCQELN 453
 QY 206 VPVFDFPHFYANP---DDHADNLVALPRMTKWTGRIGLQPKVHLSSPKSEQAI----- 256
 DB 454 IPLVLDWHHNNVPGTLRGSLDMLPLTITRETWRKGTOKOHYESADPTAISGMKR 513
 QY 257 RSHADYVDANFLERFRQWGTNIDFMIEAKOKDKALLRL 295
 DB 514 RAHSDRV-----FDFPPCDPTMDLMEAKEKEQAVFEL 546

RESULT 5
 Q01408 PRELIMINARY; PRT; 656 AA.
 ID Q01408
 AC Q01408
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE UV-ENONUCLEASE.
 GN UVEL.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=74-OR23-1A;
 RX MEDLINE=9522980; PubMed=7774597;
 RA Yajima H., Takao M., Yasuhira S., Zhao J.H., Ishii C., Inoue H.,
 RA Yasui A.;
 RT "A eukaryotic gene encoding an endonuclease that specifically repairs
 RT DNA damaged by ultraviolet light.";
 RL EMBO J. 14:2393-2399(1995).
 DR EMBL; D11392; CAB20113.1;
 KW Endonuclease.
 SQ SEQUENCE 656 AA; 74413 MW; 4B231E7A8735C86F CRC64;

Query Match 20.8%; Score 345; DB 3; Length 656;

Best Local Similarity 32.3%; Pred. No. 4.9e-20;
 Matches 87; Conservative 51; Mismatches 105; Indels 26; Gaps 8;

QY 46 ANLENTMRTLHYIIGHGIPLYRFSSSIVPLATHPDVMMDFVTFPOKE-FREIGELVYKTHQ 104
 DB 277 ANARDIVKMLCWNKEKYIRFLRLSSEMPFASHP-VHCYKLAPFASVLAEGRAVAAELG 335
 QY 105 LRTSFHPNQFTLTSPKESVTKNVDMAYHYRMLEAMGIADR-----SVINTIHGGAYGN 160
 DB 336 HRLTHPGQFTQLGSPKRVESAIRDLEYDELILLSLLKLPQONRDVAVMLIHMGQGFCD 395
 QY 161 KDTATAQFHONIKOLPOEIKERMTLENDDKTYTTEETLQVCEQEDVFPVDFHFIHYANPD 220
 DB 396 KAATLERFKRYARLSQCKNRLVLENDVGVTVHDLPLVCEELNIPMWLDYHHNHCIFD 455
 QY 221 D-----HADNLVALP---RMKTWERIGLQPKVHLSSPKSEQAI-----BSHADYVDANF 267
 DB 456 PAHLREGTLDISDPKQERIANTWKRGIKOKMHYSEP-CDGAVTPPRRRKRRPRV----- 510
 QY 268 LLERFRQWGTNIDFMIEAKOKDKALLRL 296
 DB 511 --MTLPPCPDMDLMEAKEKEQAVFEL 537

RESULT 6
 Q09509 PRELIMINARY; PRT; 305 AA.
 ID Q09509
 AC Q09509
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE UV-ENONUCLEASE.
 GN UVSCDE.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KRI;
 RA Kitayama S., Kikuchi M., Funayama T., Narumi I., Watanabe H.;
 RT "Cloning of structural gene of an alternative incision enzyme for DNA
 RT damage in Deinococcus radiodurans.";
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB033747; BAA85759.1;
 KW Endonuclease.
 SQ SEQUENCE 305 AA; 33592 MW; B94D333243E2FE4 CRC64;

Query Match 17.7%; Score 293.5; DB 2; Length 305;
 Best Local Similarity 30.1%; Pred. No. 2.9e-16;
 Matches 88; Conservative 51; Mismatches 120; Indels 33; Gaps 11;

QY 21 KTLTFARYSKLSKTERKALLTVTKANLRNMTLHYIIGHGIPLYRFSSSIVPLATHPD 80
 DB 25 RIVTLRSYRALSPAREAKLLDLYSSNKTIRGAADYCAAHDIRLYLSSSLFPMPL---D 81
 QY 81 VMWD-----FVTPFOKEFREIGELVYKTHQURTSFHPNQFTLTSPKESVTKNVDMAYH 135
 DB 82 LAGDDTGAAVLTHLAPQLLEAGHAFTDAGVRLLMHPEQFIVLNSDRPEVRESSVRAMSAH 141
 QY 136 YRMLEAMGIADRSVINIHGGAYGNKDTATAFHONIKOLPOEIKERMTLENDDKTYTTE 195
 DB 142 ARVMDGLGLA-RTPWNLL--LHGKGGRGAGLAALIPDLPVRLRLGLENDERAYS 198
 QY 196 ETLQVCEQEDVFPVDFHFIHYANP--DDHADNLVA--LPRMKTWERIGLQPK-----VHL 247
 DB 199 ELLPICEATCTPLVFDADHHVHVHDKLPDOEDPSVREWLRLARATW-----QPPEOVVHL 253
 QY 248 SSPKSEQATRSHADYVDANFLERFRQWGTNIDFMIEAKOKDKAL--LRL 296
 DB 254 SNGIEGPQDRRH-----SHLIADFPAYADVPEVEAKGKEEAIALRLM 299

[illegible]

Query Match 6.1%; Score 101.5; DB 3; Length 1690;
Best Local Similarity 20.4%; Pred. No. 17;
Matches 66; Conservative 36; Mismatches 108; Indels 113; Gaps 15;

QY 3 FRGFVNSAMSLWDASPAKTLFARYSKLSKTERKEALLTVTKANLRN-----TWRTLHY 57
DB 157 FRESSFDSNLSSTSSPPKD-----KVSVDTVSDSSTVTASMSNPTISIDLNDXHD 212
QY 58 IIGH-----GIPLRFSSSIVPLATH-----PDVMDFTVPFOKE-----FREIGE- 98
DB 213 IIKSPETPTAGLPTQKAERKASPTAIRNQAPE-SWDVKAPIKEEPHAPKIEVAEN 271
QY 99 -----LVKTHQ-----LRTSFHPNQFTLTSPKESVTNA 128
DB 272 DVAIDNVLEKRLPVLYGTHQVPHVNTSKDIKSSHIIRVFKEDNTFTTLCPLETTTS-- 329
QY 129 VTDMAHYRMLEAMGLADR-----SVINIHIGGAYGNKDTATQAQFHNKQLPOEIKERM 183
DB 330 -----ELLAIVOKKFFLESTTNQLSVCIGCVKVLDEFEK-----PLKIOMGL 373
QY 184 TLENDKTYTTELTQVCEODVPFDFHHFYANPDHDLNVALPRMIKWTWIGLQ 243
DB 374 LLLSG---YTEEDKRLMGLREDLSFVCKF-----VVENIFLRS 408
QY 244 KVLSSPKSEQAIRSHADYVDAN 266
DB 409 LTH-----DEVLISR-NYVDVN 425

RESULT 14
P95930 PRELIMINARY; PRT; 283 AA.
AC P95930;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ORF C01032.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=DSM 1617 / P2;
RX MEDLINE=97055432; PubMed=8899719;
RA Sensen C.W., Klenk H.P., Singh R.K., Allard G., Chan C.C.-Y.,
Liu Q.Y., Penny S.L., Young F., Schenk M.E., Gaasterland T.,
Doolittle W.F., Ragan M.A., Charlebois R.L.;
RT "Organizational characteristics and information content of an archaeal
genome: 156 kb of sequence from Sulfolobus solfataricus P2.";
RL Mol. Microbiol. 22:175-191(1996).
DR EMBL; Y08256; CAA69434.1;
DR InterPro; IPR001719; AP_endonuclease_2.
DR Pfam; PF01261; AP_endonuclease2; 1.
DR SMART; SM00518; AP2ec; 1.
DR PROSITE; PS00729; AP_NUCLEASE_F2.1; UNKNOWN_1.
SQ SEQUENCE 283 AA; 32015 MW; DBFB9C13F0B5975 CRC64;

Query Match 6.1%; Score 101; DB 1; Length 283;
Best Local Similarity 21.6%; Pred. No. 1.6;
Matches 52; Conservative 48; Mismatches 87; Indels 54; Gaps 13;

QY 95 EIGELVTKHQLRTSFH-PNQFTLTSPKESV-TKNAVTDMAYHYRMLEAMGLADRSVIN 151
DB 59 ETGQVARELGVRLSVHAPYFYNLCSEKEKLEASKQRILOTDRAELMGADATIA----- 112
QY 152 IHIGGAYGNKDTATQAQFHNKQLPOEIKERM-----TLENDKTYTTELT 198
DB 113 IHI--AFYKWT-PECYQNVKEGLEVIDKAREMGRNVKFGVETMAKETAFGTLDVI 169
QY 199 QVCEQED--VPFVDFHHFYANPDHDLNVALPRMIKWTWIGLQPKVHLSSPKSEQAI 256
DB 170 SISKEIKGVIPYI-DWAHTFARQGGIDYKGKIIDRLIK---ELGL---THINS-HFESLV 221

QY 257 RSHADYVD-----ANFLLEFRQWGTNIDFIEAKOKDKALLRLMDELSS 301
DB 222 YRRGKYVDEHPIIDANAPPPPEPLAKELKR-----DISTLICESPELDRALKKKEVL 277
QY 302 I 302
DB 278 L 278

RESULT 15
Q9KB81 PRELIMINARY; PRT; 823 AA.
ID Q9KB81;
AC Q9KB81;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE BH2047 PROTEIN.
GN BH2047.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001514; BAB05766.1;
KW Complete proteome.
SQ SEQUENCE 823 AA; 97055 MW; A661C8566FE8EDC7 CMC64;

Query Match 6.0%; Score 100; DB 2; Length 823;
Best Local Similarity 23.1%; Pred. No. 8.5;
Matches 75; Conservative 44; Mismatches 101; Indels 104; Gaps 18;

QY 30 KLSKTERK-----EALLIVTKANLRNMTLHYIIGHGIPHYRFSSSIVPLATHPDV-MW 83
DB 332 KLRLEKXKLLIETPLSTPEDSLNELRS--WILK-----IKDVEHPDLSLW 377
QY 84 DFTVPQKEFREIGELVKTQLRTSFH--PNQFTLTSPKESVTKNNAVTDMAHYRMLEA 141
DB 378 QERLEWLWKVQAYPYMTLGQKQELFKLESTFTETWGTGQP-----RRLGG 423
QY 142 MGIADRSVI-----NIHIGGAYGN-----KDTATAQFH--QNIQK--LPQETK 180
DB 424 EIVADRNLLYECHGPLNNIKGSIKHLVKQDVPKWLKSIKAKHGQRKQEQALAEIF 483
QY 181 ERMTELENDKTYTTELTQVCEODVPFDFHHFYANPDHDLNVALPRMIKWTWIG 240
DB 484 KMN-----YPNEDS-----VPFLKFVHDLNSHDPVH-----TWKRW 515
QY 241 LQPKVHL-----SSPKSEQAIRSHADYVDANFLLEFRQWGTNIDFIEAKOKDKA 291
DB 516 QSIKTEIEAITHAVASNPKS--VVHLSLDYQD-----FDRDMWLITSPDLMI-AKKDDSS 568
QY 292 LLRLMDELSSIRGVKRIKGGALOW 315
DB 569 QYVILGE---IHDTINWGWALQF 589

Search completed: January 15, 2002, 14:06:28
Job time: 967 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:51:16 ; Search time 57.72 Seconds
(without alignments)
226.125 Million cell updates/sec

Title: US-09-724-296-38
Perfect score: 3055
Sequence: 1 MCTTGLSLSLGRGAAPTV.....HKLKHYHDWLWENHEKLSLS 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	146.5	4.8	1579	3	US-08-755-587-184
2	145	4.7	3418	2	US-08-639-501-2
3	145	4.7	3418	2	US-08-603-753D-4
4	145	4.7	3418	3	US-09-044-946-2
5	145	4.7	3418	3	US-08-755-587-44
6	145	4.7	3418	3	US-09-044-908-2
7	145	4.7	3418	4	US-09-099-753-4
8	145	4.7	3418	4	US-08-986-106-4
9	144	4.7	2329	3	US-08-755-587-16
10	131	4.3	754	4	US-09-214-564A-2
11	129	4.2	1683	3	US-08-755-587-183
12	117	3.8	1471	3	US-08-755-587-188
13	116.5	3.8	885	1	US-08-484-105-14
14	116.5	3.8	885	1	US-08-484-106-14
15	116.5	3.8	1535	3	US-08-755-587-185
16	116.5	3.8	2343	4	US-09-324-867-2
17	115	3.8	1013	4	US-09-415-522-8
18	114.5	3.7	1589	3	US-08-755-587-189
19	113.5	3.7	1164	4	US-08-923-992A-2
20	111.5	3.6	1128	4	US-08-923-992A-6
21	111	3.6	1005	1	US-08-089-986-3
22	111	3.6	1005	1	US-08-478-585-3
23	111	3.6	1005	1	US-08-717-312-3
24	111	3.6	1005	2	US-08-266-408-3
25	111	3.6	1005	5	PCT-US94-07886-3
26	108	3.5	926	3	US-08-755-587-187
27	106	3.5	976	4	US-09-104-324B-4

28	106	3.5	1780	1	US-08-769-309A-5	Sequence 5, Appli
29	106	3.5	1780	3	US-08-994-570-5	Sequence 5, Appli
30	105.5	3.5	1098	4	US-08-923-992A-8	Sequence 8, Appli
31	105	3.4	1494	3	US-08-755-587-186	Sequence 186, App
32	105	3.4	1863	1	US-08-425-061-16	Sequence 16, Appli
33	105	3.4	1863	2	US-08-825-886-16	Sequence 16, Appli
34	104	3.1	1863	2	US-08-603-753D-2	Sequence 2, Appli
35	104	3.4	1863	4	US-09-099-753-2	Sequence 2, Appli
36	104	3.4	1863	4	US-08-986-106-2	Sequence 2, Appli
37	104	3.4	2465	2	US-08-596-291-3	Sequence 3, Appli
38	104	3.4	2465	3	US-09-100-804-3	Sequence 3, Appli
39	103.5	3.4	914	1	US-08-484-105-2	Sequence 2, Appli
40	103.5	3.4	914	1	US-08-484-106-2	Sequence 2, Appli
41	103.5	3.4	1075	2	US-08-993-228-19	Sequence 19, Appli
42	103.5	3.4	1435	2	US-08-568-459A-4	Sequence 4, Appli
43	103.5	3.4	1435	2	US-08-487-826B-4	Sequence 4, Appli
44	103	3.4	2482	1	US-08-328-254-6	Sequence 6, Appli
45	103	3.4	3248	1	US-08-353-700-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-755-587-184
; Sequence 184, Application US/08755587
; Patent No. 6045957
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Ashworth, Alan
; APPLICANT: Stratton, Michael R
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
; TITLE OF INVENTION: susceptibility gene and uses thereof.
; NUMBER OF SEQUENCES: 222
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION NUMBER: US/08/755,587
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 184:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1579 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-755-587-184

Query Match 4.8%; Score 146.5; DB 3; Length 1579;
Best Local Similarity 21.1%; Pred. No. 5.8e-05;

Matches 94; Conservative 67; Mismatches 170; Indels 115; Gaps 19;

QY 36 KEDVAMELERVGEDEQMIKRSSECNPLLOEPIASQAQFATAGTECRKSVPCGWVVK 95
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 306 KSDISLNDIKPEKNNDYMNKAG-----LLGPISNHSFGSFRASKEIKLSEHNK 360
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
QY 96 QRLFGKTAGRDVYFISPOGLKRSKSLANLHKNGETSLKPEDFDFTVLSKRGKIKRY 155
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 361 SKMFFK-----DIEEQYPTSL--ACVEIVNTLALDNOKKLSKPQIN-TVSAHLQSSVV 411
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
QY 156 KDCSMAALTSHL--QNOSSNNWNLRKTRSKCKKDVMPSPSSSELOESRGLSNFTSTHLL 213
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 412 SDCKNSHITPQMLFSKQDFNSHNH-----TPSQKAEITE-----LSTIL- 451
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
QY 214 LKDEGVDVDFNFKVRK-----PKGKVILK----- 239
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 452 ---EESGSOFEFTQPKPSYILQKSTFEVPEQNMILKTTSECRDADLHVIMNAPSTQV 508
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
QY 240 -----GIPK-----TKGCRKSCSGFVQSDSKRESVCNKADAESPEVPAQKSLD- 285
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 509 DSSKQFEGTVEIKRFAGLLKNDCKNSASGYLTDE-----NEVGRGFYSAHGTCLNV 561
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
QY 286 RTVCISDAGACGETSVTSEENSLVKKRSLSGSGNFCSEQKTSIINKFCSAKDSEHN 345
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 562 STEALQKAVKLFSDIENISEETS--AEVHPISLSSSKCHDSVWSMFKIENHNKTSSEK 618
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
QY 346 EKYEDTFLESEI--CTKVEVVERKEHLHTDILKRGSEMDNN--CSPTRK---DFTGEK 397
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 619 NKCOLLONNIEMTGTTFVEI-----TENYKRNTENEDNKYTAASRSHNLEDFGSD 671
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
QY 398 IFQEDTPIPTQIRKRTSLYFSSKYN 423
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 672 SSKNDTV---CIHKDETDLFTDOHN 694
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |

RESULT 2
US-08-639-501-2
; Sequence 2, Application US/08639501
; Patent No. 5837492
; GENERAL INFORMATION:
; APPLICANT: Tavtligian, Sean V.
; APPLICANT: Kamb, Alexander
; APPLICANT: Simard, Jacques
; APPLICANT: Couch, Fergus
; APPLICANT: Rommens, Johanna
; APPLICANT: Weber, Barbara
; TITLE OF INVENTION: Chromosome 13-linked Breast Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti
; STREET: 1201 New York Avenue N.W., Suite 1001
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 22204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/639,501
; FILING DATE: 29-APR-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/585,391
; FILING DATE: 11-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/576,559
; FILING DATE: 21-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/575,359

Query Match 4.7%; Score 145; DB 2; Length 3418;
Best Local Similarity 20.9%; Pred. No. 0.0003;
Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;

QY 36 KEDVAMELERVGEDEQMIKRSSECNPLLOEPIASQAQFATAGTECRKSVPCGWVVK 95
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 972 KSDISLNDIKPEKNNDYMNKAG-----LLGPISNHSFGSFRASKEIKLSEHNK 1026
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
QY 96 QRLFGKTAGRDVYFISPOGLKRSKSLANLHKNGETSL-KPEDFDFTVLSKRGKIKSR 154
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 1027 SKMFFK-----DIEEQYPTSL--ACVEIVNTLALDNOKKLSKPQIN-TVSAHLQSSVV 1077
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
QY 155 YKDCSMAALTSHL--QNOSSNNWNLRKTRSKCKKDVMPSPSSSELOESRGLSNFTSTHLL 212
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 1078 VSDCKNSHITPQMLFSKQDFNSHNH-----TPSQKAEITE-----LSTIL 1118
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
QY 213 LKDEGVDVDFNFKVRK-----PKGKVILK----- 239
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 1119 ---EESGSOFEFTQPKPSYILQKSTFEVPEQNMILKTTSECRDADLHVIMNAPSIG 1174
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
QY 240 -----GIPK-----TKGCRKSCSGFVQSDSKRESVCNKADAESPEVPAQKSQL 284
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 1175 QVDSKQFEGTVEIKRFAGLLKNDCKNSASGYLTDE-----NEVGRGFYSAHGTCL 1227
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
QY 285 D-RTVCISDAGACGETSVTSEENSLVKKRSLSGSGNFCSEQKTSIINKFCSAKDSE 343
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 1228 NVSTEALQKAVKLFSDIENISEETS--AEVHPISLSSSKCHD---SVVSMF---K 1277
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
QY 344 HNEKYEDTFLESEI-----IGTKVEVVERKEHLHTDILKRGSEMDNN--CSP 388
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 1278 HNDK---TVSEKNNKCOLLONNIEMTGTTFVEI-----TENYKRNTENEDNKYTA 1327
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
QY 389 TRK---DFTGEKIFQEDTPIPTQIRKRTSLYFSSKYN 423
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |
Db 1328 SRNSHNLEFDGSDSSKNDTV---CIHKDETDLFTDOHN 1363
| | | | | : : : : : | | | | : : : : : | | | | : : : : : |

RESULT 3
US-08-603-7530-4
; Sequence 4, Application US/086037530
; Patent No. 5891857
; GENERAL INFORMATION:
; APPLICANT: HOLT, JEFFREY T.
; APPLICANT: JENSEN, ROY A.
; APPLICANT: PAGE, DAVID L.
; APPLICANT: KING, MARY-CLAIRE
; APPLICANT: SZABO, CSILLA I.
; APPLICANT: JETTON, THOMAS L.
; APPLICANT: ROBINSON-BENION, CHERYL L.
; APPLICANT: THOMPSON, MARILYN E.
; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
; TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.

NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARLES A. TAYLOR, JR.
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
STREET: BOULEVARD
CITY: DURHAM
STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27707
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
COMPUTER: IBM PC/XT/AT compatible
OPERATING SYSTEM: Windows 3.1
SOFTWARE: WORD PERFECT 6.1 and ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/603,753D
FILING DATE: 20 FEB 1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/373,799
FILING DATE: 17 JAN 1995
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: female breast
CELL TYPE: normal breast tissue
CELL LINE: HMEC
ORGANELLE: no
FEATURE:

NAME/KEY: BRCA2 protein
LOCATION: 1 to 3418; Genbank locus HSU43746
IDENTIFICATION METHOD:
OTHER INFORMATION: BRCA2 protein has a negative regulatory effect on growth of human mammary cells.
PUBLICATION INFORMATION:
AUTHORS: Wooster, R. et al.
TITLE: Identification of the breast cancer
JOURNAL: Nature
VOLUME: 379
PAGES: 789-792
DATE: 1995
RELEVANT RESIDUES IN SEQ ID NO: 4: granin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344
US-08-603-753D-4

Query Match 4.7%; Score 145; DB 2; Length 3418;
Best Local Similarity 20.9%; Pred. No. 0.0003;
Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;
QY 36 KEDVAMELVERGDEEQMKKSSCECNPLQEPASOFGATAGTECKRSVPCGWVRVK 95
DB 972 KSDLSLNIDKIPKKNNDYKNKAG-----LLGPISNHSFGGSFRTASNKEIKLSEHNIK 1026

QY 96 ORLFGKTAGRFVYFISPOCLKFRSKSSLANYLHKNGETSL-KPEDFDFTVLSKRGIKSR 154
DB 1027 SKMEFK-----DIEEYPTSL---ACVEIVNTLALDNQKLLSKPOSIN-TVSAHLQSSV 1077
QY 155 YKDCSMAALTSHL--ONQSNNSNWNLRKCKKDVFPMPSSSSSELSQESRGLSNFTSTHL 212
DB 1078 VSDCKNSHITPQMLFSKQDFNSHNHL-----TPSOKAEITE-----LSTIL 1118
QY 213 LLKEDEGVDVDFRVRK-----PKGKVTILK----- 239
DB 1119 -----EESGQFEFTQPRKPSYILOKSTFEVPEENOMTILKTTSECRDADLHVINNAPSIG 1174
QY 240 -----GPIPKK-----TKGCRKSCGFGVSDSKRESVCNKADAESEPEVAQKSQL 284
DB 1175 QVDSKQFECTVEIKRKFPAGLLKNCNKSASGYLTDE-----NEVGRFGYSAHGFKL 1227
QY 285 D-RTVCISDAGACGETLSVTSEENSLVRKKERSLSSGNCFCSEOKTSGIINKFCSAKDS 343
DB 1228 NVSTEALQKAVKLFSDIENISEET'S---AEVHPISLSSSKCHD---SVVSMF---KIE 1277
QY 344 HNEKYEDTFLESEE-----IGTKVEVVERKEHLHTDILKRGSEMDNN--CSP 388
DB 1278 HNDK---TVSEKNKNCQILLQNNIMTTGTVEEI-----TENYKRNTEENEDNKYTAA 1327
QY 389 TRK-----DFTGKIFQEDTIPRTQIERKRTSLYFSSKYN 423
DB 1328 SRNSHNLEFDGSDSKNDTV---CIHKDETDLFTDOHN 1363
RESULT 4
US-09-044-946-2
Sequence 2, Application US/09044946
Patent No. 6033857
GENERAL INFORMATION:
APPLICANT: Tavtigan, Sean V.
APPLICANT: Kamb, Alexander
APPLICANT: Simard, Jacques
APPLICANT: Couch, Fergus
APPLICANT: Rommens, Johanna
APPLICANT: Weber, Barbara
TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
NUMBER OF SEQUENCES: 124
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti
STREET: 1201 New York Avenue N.W., Suite 1001
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 22204
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/044,946
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/639,501
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/576,559
FILING DATE: 21-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/575,359
FILING DATE: 20-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/573,779
FILING DATE: 18-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.

Tue Jan 15 14:12:16 2002

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/755,587

FILING DATE: 25-NOV-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9523959.6

FILING DATE: 23-NOV-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9525555.0

FILING DATE: 14-DEC-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9617961.9

FILING DATE: 28-AUG-1996

AUTHOR/AGENT INFORMATION:

NAME: Kenneth D Sibley

REGISTRATION NUMBER: 31,665

REFERENCE/DOCKET NUMBER: 5405-135

INFORMATION FOR SEQ ID NO: 44:

SEQUENCE CHARACTERISTICS:

LENGTH: 3418 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-755-587-44

Query Match 4.7%; Score 145; DB 3; Length 3418;

Best Local Similarity 20.9%; Pred. No. 0.0003;

Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;

QY 36 KEDVAMELERVCEDEEOMIKRSECNPLLOEPIASAOFGATAGTCRKSVPCCGWERVVK 95

DB 972 KSDISLNIDKIPEKNDYMNKAG-----LLGPISNHSFGSGFRTASNKEIKLSEHNKK 1026

QY 96 QRLFGKTAGRFVDYFISPOGLKFRKSSSLANYLHKNGETSL-KPEDDFDTVLSKRGIKSR 154

DB 1027 SKMFFK-----DIEEQYPTSL---ACVEIVNTLALDNOKLKSQPSIN-TVSAHLQSSVV 1077

QY 155 YKDCSMAALTSHL--ONQSNNSNNWLRTRSKCKDVFPMPSSSELOESRGLSNFTSTHL 212

DB 1078 VSDCKNSHITPQMLFSKQDFNSHNHL-----TPSQKAEITE-----LSTIL 1118

QY 213 LLKEDGVDVDFNFRKVRK-----PKGKVTILK----- 239

DB 1119 -----ESGSGFETQPRKPSYILQKSTFVPEPQNTILKTTSECRDADLHVIMNAPSIG 1174

QY 240 -----GIPDKK-----TKGCRKSCSGFVSDSKRESVCNKADAEPVAKQSOL 284

DB 1175 QVDSKQFEGTVEIKRKFAGLKNDCKNSASGYLTDE-----NEVGRFGYSAHGTKL 1227

QY 285 D-RVTCISDAGACGETLSVTSENSLVKKKRSLSGSGNFCSEQKTSIGIINKFCSAKDSE 343

DB 1228 NVSTEALQKAVKLFSDIENISEETS---AEVHPISLSSSKCHD---SVVSMF---KLEN 1277

QY 344 HNEKYEDTFLESEE-----IGTKVEVVERKEHLHTDILKRGSEMDNN--CSP 388

DB 1278 HNDK---TVSEKNNKCOLIQNNIEMTGTFFVEI-----TENYKRNTEDEKNKYTAA 1327

QY 389 TRK---DFTCEKIFQEDTIPRTQIERRKTSILYFSSKYN 423

DB 1328 SRNSHNLEFDGSDSSKNDTV---CIHKDETDLFLFTDOHN 1363

RESULT 6

US-09-044-908-2

; Sequence 2, Application US/09044908

; Patent No. 6124104

; GENERAL INFORMATION:

; APPLICANT: Tavtigan, Sean V.

; APPLICANT: Kamb, Alexander

; APPLICANT: Simard, Jacques

; APPLICANT: Couch, Fergus

REGISTRATION NUMBER: 28,957

REFERENCE/DOCKET NUMBER: 24884-116802-04

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-962-4810

TELEFAX: 202-962-8300

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 3418 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-044-946-2

Query Match 4.7%; Score 145; DB 3; Length 3418;

Best Local Similarity 20.9%; Pred. No. 0.0003;

Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23;

QY 36 KEDVAMELERVCEDEEOMIKRSECNPLLOEPIASAOFGATAGTCRKSVPCCGWERVVK 95

DB 972 KSDISLNIDKIPEKNDYMNKAG-----LLGPISNHSFGSGFRTASNKEIKLSEHNKK 1026

QY 96 QRLFGKTAGRFVDYFISPOGLKFRKSSSLANYLHKNGETSL-KPEDDFDTVLSKRGIKSR 154

DB 1027 SKMFFK-----DIEEQYPTSL---ACVEIVNTLALDNOKLKSQPSIN-TVSAHLQSSVV 1077

QY 155 YKDCSMAALTSHL--ONQSNNSNNWLRTRSKCKDVFPMPSSSELOESRGLSNFTSTHL 212

DB 1078 VSDCKNSHITPQMLFSKQDFNSHNHL-----TPSQKAEITE-----LSTIL 1118

QY 213 LLKEDGVDVDFNFRKVRK-----PKGKVTILK----- 239

DB 1119 -----ESGSGFETQPRKPSYILQKSTFVPEPQNTILKTTSECRDADLHVIMNAPSIG 1174

QY 240 -----GIPDKK-----TKGCRKSCSGFVSDSKRESVCNKADAEPVAKQSOL 284

DB 1175 QVDSKQFEGTVEIKRKFAGLKNDCKNSASGYLTDE-----NEVGRFGYSAHGTKL 1227

QY 285 D-RVTCISDAGACGETLSVTSENSLVKKKRSLSGSGNFCSEQKTSIGIINKFCSAKDSE 343

DB 1228 NVSTEALQKAVKLFSDIENISEETS---AEVHPISLSSSKCHD---SVVSMF---KLEN 1277

QY 344 HNEKYEDTFLESEE-----IGTKVEVVERKEHLHTDILKRGSEMDNN--CSP 388

DB 1278 HNDK---TVSEKNNKCOLIQNNIEMTGTFFVEI-----TENYKRNTEDEKNKYTAA 1327

QY 389 TRK---DFTCEKIFQEDTIPRTQIERRKTSILYFSSKYN 423

DB 1328 SRNSHNLEFDGSDSSKNDTV---CIHKDETDLFLFTDOHN 1363

US-08-755-587-44

; Sequence 44, Application US/08755587

; Patent No. 6045997

; GENERAL INFORMATION:

; APPLICANT: Futreal, Phillip A

; APPLICANT: Wooster, Richard F

; APPLICANT: Ashworth, Alan

; APPLICANT: Stratton, Michael R

; TITLE OF INVENTION: Materials and methods relating to the

; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer

; TITLE OF INVENTION: susceptibility gene and uses thereof.

; NUMBER OF SEQUENCES: 222

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Bell Seltzer Park & Gibson

; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107

; CITY: Raleigh

; STATE: NC

; COUNTRY: USA

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

APPLICANT: Rommens, Johanna
 APPLICANT: Weber, Barbara
 TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
 NUMBER OF INVENTION: Susceptibility Gene
 NUMBER OF SEQUENCES: 124
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Venable, Baetjer, Howard & Civiletti
 STREET: 1201 New York Avenue N.W., Suite 1001
 CITY: Washington
 STATE: DC
 COUNTRY: USA
 ZIP: 22204
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/044,908
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/639,501
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/576,559
 FILING DATE: 21-DEC-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/575,359
 FILING DATE: 20-DEC-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/573,779
 FILING DATE: 18-DEC-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Ihnen, Jeffrey L.
 REGISTRATION NUMBER: 28,957
 REFERENCE/DOCKET NUMBER: 24884-116802-04
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-962-4810
 TELEFAX: 202-962-8300
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3418 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-044-908-2

Query Match 4.7%; Score 145; DB 3; Length 3418;
 Best Local Similarity 20.9%; Pred No. 0.0003;
 Matches 96; Conservative 67; Mismatches 158; Indels 138; Gaps 23:
 QY 36 KEDVAMELERVGEDEOMIKRSECNPILOEPIASQAGATAGTECRKSVPCGWERVVK 95
 Db 972 KSDLSINIDKIPEKNNDYMNKAG-----LGPISNHSFGGSRFTASNKELSHNIKK 1026
 QY 96 QRLFGKTAGRFDYFTSPQGLFKRSKSLANLHKNGETSL-KPEDDFDTLSKRGIKSR 154
 Db 1027 SKWFFK-----DIEEYPTSL---ACVEIVNTALDNDOKLKKPSIN-TVSAHLQSSV 1077
 QY 155 YKDCSMAALTSL--QNSNNSNWLRTSKCKDVMPSPSSSELOESRGLSNFTSTHL 212
 Db 1078 VSDCKNSHITPQMLFSKQDFNSNHL-----TPSQKAEITE-----LSTIL 1118
 QY 213 LLKEDGVDVNVKVRK-----PKGKVTILK----- 239
 Db 1119 -----EESGQFETFRKPSYILQKSTFEVPEQNMTILKTTSECRDADLHVINNPSIG 1174
 QY 240 -----GPIKK-----TKGCRKSCGFGVQSDSKRESVCNKADASEPVAQKSQL 284
 Db 1175 QVDSSKQFEGTVEIKRKAGLLKNDCKNSASGYLDE-----NEVGFRGFYSAHGTKL 1227

QY 285 D-RTVCISDAGACGETLSVTSEENSILVKKKRSLSGSGNFCSEQKTSGLINKFCSAKDSE 343
 Db 1228 NVSTEALQKAVKLFSDIENISEETS---AEVHIPISLSKSKCHD---SVVSMF---KLEN 1277
 QY 344 HNEKYEDTFLSEEE-----IGTKVEVVVERKEHLHTDILKRGSEMDNN--CSP 388
 Db 1278 HNDK---TVSEKNNKCOLLONNIEMTTCTFVEEI-----TENYKRNTEDEDKYTA 1327
 QY 389 TRK----DFTGEKIFQEDTIPRTOHERRKTSLSYFSSKYN 423
 Db 1328 SRSHNLEFDGSDSSKNQTV---CIHKDETLLFTDOHN 1363
 RESULT 7
 US-09-099-753-4
 : Sequence 4, Application US/09099753
 : Patent No. 6149903
 : GENERAL INFORMATION:
 : APPLICANT: HOLT, JEFFREY T.
 : APPLICANT: JENSEN, ROY A.
 : APPLICANT: PAGE, DAVID L.
 : APPLICANT: KING, MARY-CLAIRE
 : APPLICANT: SZABO, CSILLA I.
 : APPLICANT: JETTON, THOMAS L.
 : APPLICANT: ROBINSON-BENION, CHERYL L.
 : APPLICANT: THOMPSON, MARILYN E.
 : TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
 : TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
 : TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
 : NUMBER OF SEQUENCES: 29
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: ARLES A. TAYLOR, JR.
 : STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
 : STREET: BOULEVARD
 : CITY: DURHAM
 : STATE: NORTH CAROLINA
 : COUNTRY: USA
 : ZIP: 27707
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette, 3.50 inch, 800 KB storage
 : COMPUTER: IBM PC/XT/AT compatible
 : OPERATING SYSTEM: Windows 3.1
 : SOFTWARE: WORD PERFECT 6.1 and ASCII
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/099,753
 : FILING DATE:
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/603,753
 : FILING DATE: 20 FEB 1996
 : APPLICATION NUMBER: U.S. 08/373,799
 : FILING DATE: 17 JAN 1995
 : ATTORNEY/AGENT INFORMATION:
 : NAME: ARLES A. TAYLOR, JR.
 : REGISTRATION NUMBER: 39,395
 : REFERENCE/DOCKET NUMBER: 1242/2
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (919) 493-8000
 : TELEFAX: (919) 419-0383
 : TELEX:
 : INFORMATION FOR SEQ ID NO: 4:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 3418
 : TYPE: amino acid
 : STRANDEDNESS: single
 : TOPOLOGY: unknown
 : MOLECULE TYPE: protein
 : HYPOTHETICAL: no
 : ANTI-SENSE: no
 : ORIGINAL SOURCE:
 : ORGANISM: Homo sapiens sapiens
 : INDIVIDUAL ISOLATE:
 : DEVELOPMENTAL STAGE: adult

us-09-724-296-38.ra1

Tue Jan 15 14:12:16 2002

TISSUE TYPE: female breast
CELL TYPE: normal breast tissue
CELL LINE: HMEC
ORGANELLE: no
FEATURE:
NAME/KEY: BRCA2 protein
LOCATION: 1 to 3418; Genbank locus HSU43746
IDENTIFICATION METHOD:
OTHER INFORMATION: BRCA2 protein has a negative
PUBLICATION INFORMATION:
AUTHORS: Wooster, R. et al.
TITLE: Identification of the breast cancer
TITLE: susceptibility gene BRCA2
JOURNAL: Nature
VOLUME: 379
PAGES: 789-792
DATE: 1995
RELEVANT RESIDUES IN SEQ ID NO: 4: granin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344
US-09-099-753-4

Query Match 4.7%; Score 145; DB 4; Length 3418;
Best Local Similarity 20.9%; Pred. No. 0.0003;
Matches 96; Conservative 67; Mismatches 138; Gaps 23;
QY 36 KEDVAMELERGVGEDEOMIKRSSECNPLLOEPDIAAQAQATAGTECRKSVPCGWVVK 95
DB 972 KSDLSLMDIKPIKNDYMNKAG-----LLGPISSHSGSGSFTASNKKEIKLSEHNKK 1026
QY 96 QRLFGKTAGFDVYFISPOGLKFRKSSSLANYLHKNGETSL-KPEDFDFTVLSKRGIKSR 154
DB 1027 SKMPEK-----DIEEQYPTSL---ACVEIVNTLALDQKLLSKPOSIN-TVSAHLQSSVV 1077
QY 155 YKDCSMAALTSHL--QNSNNNNLRTSKCKKDVMPSPSSSELOESRGLSNFTSTHL 212
DB 1078 VSDCKNSHITPQMLFSKQDFNSHNL-----TPSKAEITE-----LSTIL 1118
QY 213 LLKEDEGVDDVNFVRK-----PKGKVTILK----- 239
DB 1119 -----EESGQFEFTQFRKPSYILQKSTFEVPEMNQILKTTSECDADLHVMNAPSIG 1174
QY 240 -----GIPIKK-----TKGCRKSCSGFVQSDSKRESVCKNKADESEPVAKQSOL 284
DB 1175 QVDSKQFEGVEIKRFAAGLLKNDCKNSASGYLTDE-----NEVGRGFYSAGTKL 1227
QY 285 D-RWVCISDAGACGETLSVTSEENSLVKKERSLSSGNSFCPSQKTSGLIINKFCSAKDSE 343
DB 1228 NVSTALQKAVKLFSDIENISEETS---AEVHPISLSSSKCHD---SVVSMF---KIEN 1277
QY 344 HNEKYEDFLESEE-----IGTKVEVVEKKEHLHTDILKRGSEMDNN--CSP 388
DB 1278 HNDK---TVSEKNNKQLILONNIEMTGTGFVEI-----TENYKRNEDNEDNYTAA 1327
QY 389 TRK---DFTGEKFOEDTIPRTOIERRTKTSLYFSSKYN 423
DB 1328 SRNSHNLFDGSDSKNDTV---CLHDETDLLFTDQHN 1363

RESULT 8
US-08-986-106-4
Sequence 4, Application US/08986106
Patent No. 6177410
GENERAL INFORMATION:
APPLICANT: HOLT, JEFFREY T.
APPLICANT: JENSEN, ROY A.
APPLICANT: KING, MARY-CLAIRE
APPLICANT: STEINER, MITCHELL S.
APPLICANT: ROBINSON-BENION, CHERYL L.
APPLICANT: THOMPSON, MARILYN E.
TITLE OF INVENTION: THERAPEUTIC METHODS FOR
TITLE OF INVENTION: PROSTATE CANCER

NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARLIS A. TAYLOR, JR.
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
STREET: BOULEVARD
CITY: DURHAM
STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27707
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: IBM PC/XT/AT compatible
OPERATING SYSTEM: Windows 3.1
SOFTWARE: WORD PERFECT 6.1 and ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/986,106
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/603,753
FILING DATE: 20 FEB 1996
ATTORNEY/AGENT INFORMATION:
NAME: ARLIS A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
FEATURE:
NAME/KEY: BRCA2 protein
PUBLICATION INFORMATION:
AUTHORS: Wooster, R. et al.
TITLE: Identification of the breast cancer
TITLE: susceptibility gene BRCA2
JOURNAL: Nature
VOLUME: 379
PAGES: 789-792
DATE: 1995
RELEVANT RESIDUES IN SEQ ID NO: 4: granin box domain at
RELEVANT RESIDUES IN SEQ ID NO: amino acids 3334-3344
US-08-986-106-4

Query Match 4.7%; Score 145; DB 4; Length 3418;
Best Local Similarity 20.9%; Pred. No. 0.0003;
Matches 96; Conservative 67; Mismatches 138; Gaps 23;
QY 36 KEDVAMELERGVGEDEOMIKRSSECNPLLOEPDIAAQAQATAGTECRKSVPCGWVVK 95
DB 972 KSDLSLMDIKPIKNDYMNKAG-----LLGPISSHSGSGSFTASNKKEIKLSEHNKK 1026
QY 96 QRLFGKTAGFDVYFISPOGLKFRKSSSLANYLHKNGETSL-KPEDFDFTVLSKRGIKSR 154
DB 1027 SKMPEK-----DIEEQYPTSL---ACVEIVNTLALDQKLLSKPOSIN-TVSAHLQSSVV 1077
QY 155 YKDCSMAALTSHL--QNSNNNNLRTSKCKKDVMPSPSSSELOESRGLSNFTSTHL 212
DB 1078 VSDCKNSHITPQMLFSKQDFNSHNL-----TPSKAEITE-----LSTIL 1118
QY 213 LLKEDEGVDDVNFVRK-----PKGKVTILK----- 239
DB 1119 -----EESGQFEFTQFRKPSYILQKSTFEVPEMNQILKTTSECDADLHVMNAPSIG 1174
QY 240 -----GIPIKK-----TKGCRKSCSGFVQSDSKRESVCKNKADESEPVAKQSOL 284
DB 1175 QVDSKQFEGVEIKRFAAGLLKNDCKNSASGYLTDE-----NEVGRGFYSAGTKL 1227

QY 285 D-RTVCISDAGACGETLSVTSEENSLVKKERSLSSGNSFCSEBQKTSGLINKFCSAKDSE 343
 Db 1228 NVSTEALOKAVKLFSDIENISEETS---AEVHPISLSSSKCHD---SVVSMF---KIEN 1277
 QY 344 HNEKYEDTFLESEE-----IGTKVEVVERKEHLHTDILKRGSEMDNN--CSP 388
 Db 1278 HNDK---TVSEKNKCOLILONNIEMTCTFVEEI-----TENYKNTENEDNKYTAA 1327
 QY 389 TRK----DFTGEKIFQEDTIPRTQIERRTKTSLYFSSKYN 423
 Db 1328 SRNSHNLFFDGSDDSKNDTV---CIHKDETDLLETDQHN 1363

RESULT 9

US-08-755-587-16

; Sequence 16, Application US/08755587

; Patent No. 6045997

; GENERAL INFORMATION:

; APPLICANT: Futreal, Phillip A

; APPLICANT: Wooster, Richard F

; APPLICANT: Ashworth, Alan

; APPLICANT: Stratton, Michael R

; TITLE OF INVENTION: Materials and methods relating to the

; TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer

; TITLE OF INVENTION: susceptibility gene and uses thereof.

; NUMBER OF SEQUENCES: 222

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Bell Seltzer Park & Gibson

; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107

; CITY: Raleigh

; STATE: NC

; COUNTRY: USA

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/755,587

; FILING DATE: 25-NOV-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9523959.6

; FILING DATE: 23-NOV-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9525555.0

; FILING DATE: 14-DEC-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9617961.9

; FILING DATE: 28-AUG-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Kenneth D Sibley

; REGISTRATION NUMBER: 31,665

; REFERENCE/DOCKET NUMBER: 5405-135

; INFORMATION FOR SEQ ID NO: 16:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2329 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-755-587-16

Query Match

Best Local Similarity 21.1%; Score 144; DB 3; Length 2329;

Matches 97; Conservative 66; Mismatches 150; Indels 138; Gaps 24;

QY 36 KEDVAMELVRGDEEOMIKRSSECNPLLPPIASAOFCATAGTECRKSVPCGWERVVK 95

Db 691 KSDISLNDIKPEKNDYMKWAG-----LLGPISNHSEGGSFRTASKEIKLSHNIKK 745

QY 96 QRLFGKTAGFDVYFISPOGLFKRSKSIANYLHKNGETSL-KPEDFDFTVLSKRGIKSR 154

Db 746 SKMFFK-----DIEEQYPTSL---ACVEIVNTALDQNKLSKPKQSIN-TVSAHLQSSV 796

QY 155 YKDCSMAALTSHL--QNQSNNSNNLRTSKCKKQVMPSPSSSELQESRGLSNFTSTHL 212
 Db 797 VSCKNSHIITPOMLFKQDFNSNHL-----TPSQEQITE-----LSTIL 837
 QY 213 LLKEDGCVDDVNRKVRK-----PKGKVTILK----- 239
 Db 838 ---EDSG-SQFEFTQFRKPSYILOKSTFEVPENQMTILKTTSSECRDADLHVIMNAPSIG 893
 QY 240 -----GPIKK-----TKGCRKSCSGFVOSDKKRESVCNKADAESPEVAKKSOL 284
 Db 894 QVDSKQFEGTVEIKRKFAGLLKNCNKASGYLTDE-----NEVGRFGYSAHGTKL 946
 QY 285 D-RTVCISDAGACGETLSVTSEENSLVKKERSLSSGNSFCSEBQKTSGLINKFCSAKDSE 343
 Db 947 NVSTEALOKAVKLFSDIENISEETS---AEVHPISLSSSKCHD---SVVSMF---KIEN 996
 QY 344 HNEKYEDTFLESEE-----IGTKVEVVERKEHLHTDILKRGSEMDNN--CSP 388
 Db 997 HNDK---TVSEKNKCOLILONNIEMTCTFVEEI-----TENYKNTENEDNKYTAA 1046
 QY 389 TRK----DFTGEKIFQEDTIPRTQIERRTKTSLYFSSKYN 423
 Db 1047 SRNSHNLFFDGSDDSKNDTV---CIHKDETDLLETDQHN 1082

RESULT 10

US-09-214-564A-2

; Sequence 2, Application US/09214564A

; Patent No. 6150515

; GENERAL INFORMATION:

; APPLICANT: Sharp, Phillip A.

; APPLICANT: Zhou, Qiang

; TITLE OF INVENTION: TAT-SF: Cofactor For Stimulation Of Transcriptional

; FILE REFERENCE: M0656/7042

; CURRENT APPLICATION NUMBER: US/09/214,564A

; PRIOR FILING DATE: 1999-08-18

; PRIOR APPLICATION NUMBER: US 60/021,218

; PRIOR FILING DATE: 1996-07-03

; PRIOR APPLICATION NUMBER: US 60/033,152

; PRIOR FILING DATE: 1996-12-13

; PRIOR APPLICATION NUMBER: PCT/US97/11713

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 754

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-214-564A-2

Query Match

Best Local Similarity 20.3%; Score 131; DB 4; Length 754;

Matches 88; Conservative 69; Mismatches 161; Indels 116; Gaps 17;

QY 13 DRGNAPTVTSERLV-----PDP-----NDLRKEDVAMELERVGEDEFEQMMIKRS 58

Db 252 ERRAGPSRMHRHVITKNMFHMDPDDPLVLNEIR-EDLRVECSKFCQIRKLLFLDRH 310

QY 59 SECNPLLPPIASAOFCATAGTE-CRKSVPCGWERVVKORLFCK---TAGRFDVYFISQ 114

Db 311 -----PDGVASVSTFDEEADYCIOTLDGRW-----EGRQJITAAQWD----- 348

QY 115 GLKFRKSIANYLHKNGETSLKPEDFDFTVLSKRGIKRSYKDCSMAALTSHLQNSNS 174

Db 349 -----GTTDYQVEE-----TSREEREKLRGWEAFNLNPEANR 380

QY 175 NNWLRTSKCKKQVMPSPSSSELQD---SRGLSNFTSTHLLKEDEGVDDVNFNRK----- 227

Db 381 GLSVQIILSLLRK---AGPSRKRHFSEHPSTSKMNAQETATGMAFEPIDKKFKTEKDG 437

us-09-724-296-38.ra1

Tue Jan 15 14:12:16 2002

QY 228 -----VRKPKGKVTILKIPKTKKCKSCGFGVQSDSKRESVCNKADAESEP 277
 Db 438 EFEEGASENNAKESPEKEAEGCGPEKESGCPK--RGFEGSCQKESGEGNPNVRGSEE 495
 QY 278 VAKQSOLDRTVCISDAGAGCTLSVTSEENSLVKKERSLSSGS-----NFCSEQKTS 330
 Db 496 DSPKRESK-----KTLKNDCEENGLAKESDDLNKSEEEVGVTKSEEDDSE 544
 QY 331 GIINKFCSAKOSEHNEKYEDTFLESEETGTVKVVVER--KEHLHTDILKRGSEMDNMCSP 388
 Db 545 KESDECSKQSGEDSE-----REFEENGLKDLDEESKELHENVLDKLELENDSENS 599
 QY 389 TRKDTGTEKIFQED 402
 Db 600 EFEDDGSKEVLDEE 613

RESULT 11
 US-08-755-587-183
 ; Sequence 183, Application US/08755587
 ; Patent No. 6045997
 ; GENERAL INFORMATION:
 ; APPLICANT: Futreal, Phillip A
 ; APPLICANT: Wooster, Richard F
 ; APPLICANT: Ashworth, Alan
 ; APPLICANT: Stratton, Michael R
 ; TITLE OF INVENTION: Materials and methods relating to the
 ; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
 ; TITLE OF INVENTION: susceptibility gene and uses thereof.
 ; NUMBER OF SEQUENCES: 222
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Bell Seltzer Park & Gibson
 ; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
 ; CITY: Raleigh
 ; STATE: NC
 ; COUNTRY: USA
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/755,587
 ; FILING DATE: 25-NOV-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9523959.6
 ; FILING DATE: 23-NOV-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9525555.0
 ; FILING DATE: 14-DEC-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9617961.9
 ; FILING DATE: 28-AUG-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kenneth D Sibley
 ; REGISTRATION NUMBER: 31,665
 ; REFERENCE/DOCKET NUMBER: 5405-135
 ; INFORMATION FOR SEQ ID NO: 183:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1683 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; US-08-755-587-183

Query Match 4.28; Score 129; DB 3; Length 1683;
 Best Local Similarity 20.48; Pred No. 0.0035;
 Matches 103; Conservative 67; Mismatches 175; Indels 160; Gaps 19;
 QY 11 LGDRGAATPV-----TSSRLVPDPNDLR--KEDVAMELEKVEDEQMKRSSECN-- 62
 Db 297 IGDKQAQVSIKKDLDSIAVHDLAEENRNSKQHLKMTLDDDLKSDLSLNXKSGDND 356

QY 63 -----PLLOEPFASOFGAAGTECRKSVPCGWFRVVKQRLFKTAGREDVYFISPGQLK 117
 Db 357 YMDKAGLLDPLSNHSGSFRASNKEIKLSEHNKSKMFKDIE--EQYPTSLACVE 414
 QY 118 FRSSSLANYLHKNGETSLKPEDFTVLKSRGKISRYKDCSMAALTSHLN----- 169
 Db 415 JVTNLAN-----OKLSKXXXDUPQINTVSAHLQSSVVVSDCE 454
 QY 170 -----OSNNNSNNLRTSKCKDVFMPPSPSSSELOESRGLSNFTSTHLLKED 217
 Db 455 DSHTAPOMLSKODFNSNHL-----TPSOKAEITE-----LSTIL---E 491
 QY 218 EGVDDVNFVRK-----PRGKVTILK----- 239
 Db 492 ESGSQFETQFKRKSHIAQKNTFEPENQMTILNTSEWKDADLHLIVNAPSIOVDSSK 551
 QY 240 -----GPIPK---KTKGCRKSCGFGVQSDSKRE-----SYCNKADAESEPVAKOSOL 284
 Db 552 QFEGSAGIKQKFACLLKSSCNKSASGYLTDENEVEFRGCFYSALGTKLVNVSSEALOKAVKL 611
 QY 285 DRTVCISDAGAGCTLSVTSEENSLVKKERSLSSGSNFCSEQKTSIINKFCSAKDSEH 344
 Db 612 -----FSDIENLSEETSAEVDPISSLSSKYHD-SVASFKNENQXD-----KSSEK 657
 QY 345 NEKYEDTFLESEETGTVKVVVERKEHLHTDILKRGSEMDN-----CSPTRKDTGTEK 397
 Db 658 NKKCOLILQNNLEMTIGIFVEE-----NTENYKRNTEENEDNKYTGASRNSXNLEXSDGSD 712
 QY 398 IFQEDTIPRTQIERKRTSLYFSSKY 422
 Db 713 SSKNDTV---YHKDETDLFPIDQH 734

RESULT 12
 US-08-755-587-188
 ; Sequence 188, Application US/08755587
 ; Patent No. 6045997
 ; GENERAL INFORMATION:
 ; APPLICANT: Futreal, Phillip A
 ; APPLICANT: Wooster, Richard F
 ; APPLICANT: Ashworth, Alan
 ; APPLICANT: Stratton, Michael R
 ; TITLE OF INVENTION: Materials and methods relating to the
 ; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
 ; TITLE OF INVENTION: susceptibility gene and uses thereof.
 ; NUMBER OF SEQUENCES: 222
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Bell Seltzer Park & Gibson
 ; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
 ; CITY: Raleigh
 ; STATE: NC
 ; COUNTRY: USA
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/755,587
 ; FILING DATE: 25-NOV-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9523959.6
 ; FILING DATE: 23-NOV-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9525555.0
 ; FILING DATE: 14-DEC-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9617961.9
 ; FILING DATE: 28-AUG-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kenneth D Sibley
 ; REGISTRATION NUMBER: 31,665
 ; REFERENCE/DOCKET NUMBER: 5405-135

INFORMATION FOR SEQ ID NO: 188:
SEQUENCE CHARACTERISTICS:
LENGTH: 1471 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-755-587-188

Query Match 3.8%; Score 117; DB 3; Length 1471;
Best Local Similarity 22.4%; Pred. No. 0.043;
Matches 85; Conservative 45; Mismatches 148; Indels 102; Gaps 15;
Qy 63 PLLOEPIASQFGATAGTCERKSVPCGWERVVKORLPFGKTAGRDYVYFISPOGLKPKRSK 122
Db 297 PGFLDPVFKHKGSGSFRSTASNKELKSEHNVKSKMFFKDIE--EYPTSLTICIDIVNAS 354
Qy 123 SLANYLKHNGETSLKPDFDFTVL-----SKRGIKSRKDCGSMALTSHLQNSNN 175
Db 355 PLANOELISG-----PYTDLQVTTMSAHPQSQASVSCEDTHTSLQVLPs---KQDFHSN 406
Qy 176 WNLRTSRCKKDDVMPSPSSSELOESRGLNFTTSHLLKKEDEGVDDVNFVKRKKPK--- 232
Db 407 HNL-----TPSOKAEITE-----LSTIL-----EESGSOFEFTQFKPKSHVA 443
Qy 233 -----GKVTILKGIPIKTKKCRK-----SCSGFVQSDSKRESVCNKADAEPV 278
Db 444 QNNIPEVPGKT---VAINTTSGWKRRIGLHLTVDPASVOTDDSKRKEGSAFQSGFPC 499
Qy 279 AQKSLDRVTVCISDA-----GACGFTLSVTSE-----ENSLVK 311
Db 500 LKSSCNKNTSFLANVEMEFGRFSALGTLKLSVSEALQKAVKLPDIESGSEETSTK 559
Qy 312 KERSLSGSGNFCSEQKTSGI-----INKFCAKSEHNEKYEDTFLESEELGTQKVEVER 367
Db 560 VDPRALSSGA-----RHDGASVFKIRKQNSGKSDKTSKQVTLQNNTEVTGIFVDRN 614
Qy 368 KEHLHTDILKRGSEMDNCS 387
Db 615 PEN-----YARNTKCEDNNS 629

RESULT 13
US-08-484-105-14
Sequence 14, Application US/08484105
Patent No. 5589341
GENERAL INFORMATION:
APPLICANT: STILLMAN, Bruce
APPLICANT: BELL, Stephen P
APPLICANT: KOBAYASHI, Ryuji
APPLICANT: RINE, Jasper
APPLICANT: FOSS, Margit
APPLICANT: McNALLY, Francis J
APPLICANT: LAURENSEN, Patricia
APPLICANT: LI, Joachim J
APPLICANT: GAVIN, Kimberly
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,105
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 885 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-105-14

Query Match 3.8%; Score 116.5; DB 1; Length 885;
Best Local Similarity 20.4%; Pred. No. 0.02;
Matches 79; Conservative 65; Mismatches 141; Indels 103; Gaps 18;
Qy 222 DVNF--RKVRKPKGVTI--LKGIPIKTKKGCRCSCGFSQSDSKRESVCNKADAEPV 277
Db 192 DIHFFEEKIRSSNPKVSLVLRDITLPLKLPKPLNKS-----KVHAREKV 235
Qy 278 VAKQSOLDRTVCISDAGAGETLSVTSEE-----NSLVKKKERSLSSGNSFCSEQKTSGI 332
Db 236 VAKQLQSDNT---PSKKSEFOQTVSKTNAEVORIASIVNEKE-AISDNESDLSEYHESK- 290
Qy 333 INKFCSAKDSHNEKYEDTFLESEELGTQKVEVERKEHLHTDILKRGSEMDNCSPTRKD 392
Db 291 -EFANASSDSDEEED-YOSAEELAIPEPAKKKRSIKPDIPISPVKSQTPLOPSAVH 348
Qy 393 FTGKIFQEDTIPRTQERRKTSLYFSSKYNKALSPRRKPAKPKWTTPRSPNLYOETL 452
Db 349 SSPKRFKNNI-----VRAKKAVTPFSKRY-KNPKIPDLNDIFORHN-----NLDLIA 496
Qy 453 FHDHPKLLIA-----TIF-----LNRTSKMAIPVLKFEKYPSEAE----- 489
Db 397 LEERFRTVSAKGMETIFSKVKOLNSRKEELVKAADFNLYLPARENENFASIIYLSLYS 456
Qy 490 -----VARTADWRDVSSELLKPLGLYDLRAKTIIVKFSDEYLTQKWKYPI 532
Db 457 AIDAGTSTSIYIAGTGVGKTLTVREV-----VKDLMTSADOKELPRQY-I 502
Qy 533 ELHGIG-KYGNDSYRIFCVNENKQVHPE 559
Db 503 EINGLKIVKASDSYEVF---WQKISGE 526

RESULT 14
US-08-484-106-14
Sequence 14, Application US/08484106
Patent No. 5614618
GENERAL INFORMATION:
APPLICANT: STILLMAN, Bruce
APPLICANT: BELL, Stephen P
APPLICANT: KOBAYASHI, Ryuji
APPLICANT: RINE, Jasper
APPLICANT: FOSS, Margit
APPLICANT: McNALLY, Francis J
APPLICANT: LAURENSEN, Patricia
APPLICANT: HERSKOWITZ, Ira
APPLICANT: LI, Joachim J
APPLICANT: GAVIN, Kimberly
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco

STATE: California
 COUNTRY: USA
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/484,106
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Osman Ph.D., Richard Aron
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 494-8700
 TELEFAX: (415) 494-8771
 TELEX: 910 277299
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 885 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-484-106-14

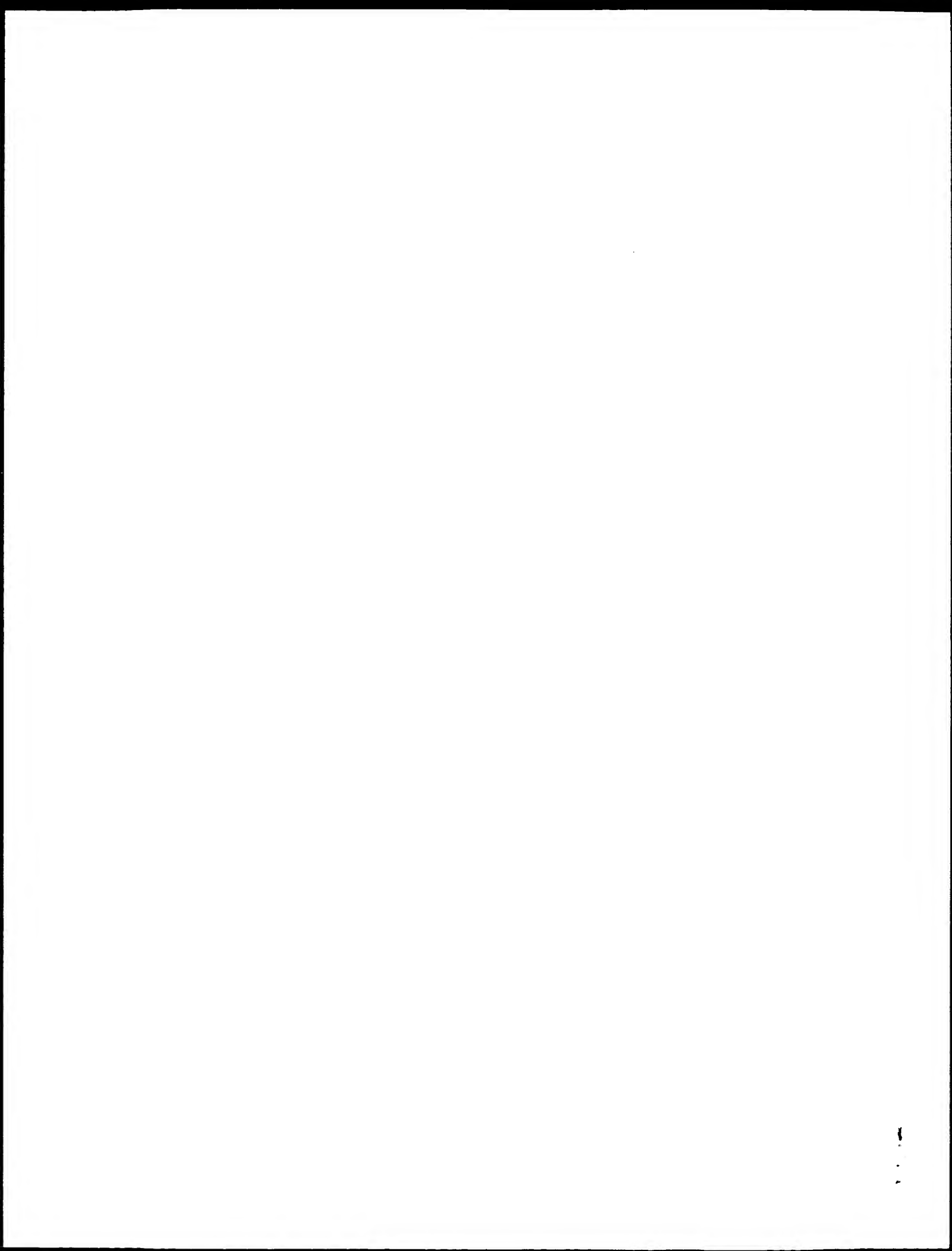
Query Match 3.8%; Score 116.5; DB 1; Length 885;
 Best Local Similarity 20.4%; Pred. No. 0.02;
 Matches 79; Conservative 65; Mismatches 141; Indels 103; Gaps 18;
 QY 222 DVNF--RKVRPKGKVTI--LKGPIKTKKCGKSCSGFVQSDSKRSVCNKADESEP 277
 Db 192 DIHFEKIRSNPKVSLVLRITLPLKLPKLNKS-----KVHAREKV 235
 QY 278 VAQKSQLDRVTCISDAGAGETLSVTSEE-----NSLVKKKERSLSSGNFCSEKTSGT 332
 Db 236 VARKLQSDNT---PSKSFQOTVSKTNAEVQRIASTIVNEKE-AISDNESDLSEYHESK- 290
 QY 333 INKFCSAKDEHNEKYEDTFLESEEIGTVKEVVERKEHLHTDILKRGSEMDNCSPTKRD 392
 Db 291 -EFANASSDSDEED-YQSAELAIYEPKAKKVRSTKPDIPISVKSQTPLOPSAVH 348
 QY 393 FTGEKIFQEDTTPRTQIERRKTSYLFSSKYNKEALSPPRKAKFKWTPPRSPFNLVQETL 452
 Db 349 SSPKFEKNNI-----VRAKKAYTPFSKRY-KNPKTPDLNDIFQRHN-----NDLDIAA 396
 QY 453 FHDPWKLLIA---TIF-----LNRTSGKMAIPVLKFKLEKYPSE-----489
 Db 397 LBERFTVSAGKMETIFSKVKQLNSRNSKEEIVKAADFNYLPARENEFASISLSYS 456
 QY 490 -----VARTADWRDVSSELLKPLGLYDLIRAKTIVKFSDEVLTKQWKYPI 532
 Db 457 AIEAGTSTSIYIAGTPGVGKTLTVREV-----VKDLMTSADQELPRFOY-I 502
 QY 533 ELHGIG-KYGNDSYRIFCVNEHKQVHPE 559
 Db 503 EINGLIVRASDSYEVF-----WOKISGE 526

RESULT 15
 US-08-755-587-185
 ; Sequence 185, Application US/08755587
 ; Patent No. 6045957
 ; GENERAL INFORMATION:
 ; APPLICANT: Futreal, Phillip A
 ; APPLICANT: Wooster, Richard F
 ; APPLICANT: Ashworth, Alan
 ; APPLICANT: Stratton, Michael R
 ; TITLE OF INVENTION: Materials and methods relating to the

TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
 TITLE OF INVENTION: susceptibility gene and uses thereof.
 NUMBER OF SEQUENCES: 222
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Bell Seltzer Park & Gibson
 STREET: 310 UCR Plaza, 3605 Glenwood Avenue, PO Drawer 31107
 CITY: Raleigh
 STATE: NC
 COUNTRY: USA
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/755,587
 FILING DATE: 25-NOV-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9523959.6
 FILING DATE: 23-NOV-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9525555.0
 FILING DATE: 14-DEC-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9617961.9
 FILING DATE: 28-AUG-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Kenneth D Sibley
 REGISTRATION NUMBER: 31,665
 REFERENCE/DOCKET NUMBER: 5405-135
 INFORMATION FOR SEQ ID NO: 185:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1535 amino acids
 TYPE: amino acid
 TOPOLOGY: Linear
 US-08-755-587-185

Query Match 3.8%; Score 116.5; DB 3; Length 1535;
 Best Local Similarity 20.4%; Pred. No. 0.051;
 Matches 97; Conservative 75; Mismatches 163; Indels 141; Gaps 26;
 QY 20 VTSERLVPPDPDLRKEDVAMELERVGEDFEQMMIKRSSECNPLLOEPIASQOFGATAG 79
 Db 293 MTLQGDQLPD-----ILNIDKIPDKIDDCMDKWA---GPL--DPISNHSFGSFR 337
 QY 80 TECPKSVPCGWVVKORLFGKTAGRDVYFISPOGLKFRKSLANYLHKNGETSLKPE 139
 Db 338 TASKEIKVSEHNKIKSKMPFK-----DIEEQYPTSL---ACVEIVNTLALDNQKLSKP 389
 QY 140 DFDFTVLSKRGKIRYKDCSMAALTSHL---QNGSNNSNNLRTSRCKKQDVFPMPSSSE 197
 Db 390 QIN-TVSAHLQSSVVSDCKNSHITPMQSFQKDFNSNHL-----TPQKAE 436
 QY 198 LOESRGLSNPTSTHLLKDEGEVDVDFNFRVKR-----PKGVTKLK-----239
 Db 437 ITE-----LSTIL-----EESGQFEFTQFRKPSYLOKNTFEVPEVNOVTLNTEEN 485
 QY 240 -----GPIPKTKK-----GCRKSCSGFVQSDSKRSVCNKADESEPAQKQ 283
 Db 486 RDAGLVIMNAPSTQVNSSKQFECTVGIKQKFAGLLKS-----CNK--SASYLTDENE 537
 QY 284 LDRVTCISDAGAGETLSVTSEENSLVKKERSLSSGNSFCSEKTSGTI-----NKF 336
 Db 538 VEPGRFYS---AHGVKLVNSTE---ALQRAVKLFSDIENI---SEKTSAEVDPISLSSSKF 589
 QY 337 CSA-----KDSEHNEKYD-----TFLESE---EIGTKVEVVERKEHLHTDILKREMD 383
 Db 590 HDSVWSMFKTENHNDKTSKNCQMLQNIEMTTTGTTFVEEITENTYKINTE-----NED 643
 QY 384 NNCSPTRK-----DFTGKIFQEDTI-----PRTOIERRKTSYFESSKYNKE 425
 Db 644 NKVTAASRNSRNLEFPVGSDDSKNDVTCLHKDKDLPFT--DQNRICKLSGQLMKE 697

Search completed: January 15, 2002, 13:51:25
Job time: 354 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:52:32 ; Search time 74.94 Seconds
(without alignments)
589.555 Million cell updates/sec

Title: US-09-724-296-38
Perfect score: 3055
Sequence: 1 MGTTLGLESLSLGRGAAPTV.....HKLNKYHDWLNENHEKLSLS 580

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	209	6.8	476	S57963	methyl CpG binding
2	201	6.6	492	A41907	methyl-CpG-binding
3	145	4.7	1301	S51323	SAC3 protein - yea
4	145	4.7	3418	G02334	breast cancer tumo
5	141.5	4.6	782	S27833	rhostry-associated
6	138.5	4.5	500	S55785	nucleolar protein
7	137	4.5	560	S53382	protein YK028c ho
8	136.5	4.5	598	B40713	cylicin I - human
9	136.5	4.5	1359	T34036	hypothetical prote
10	135	4.4	822	T41622	probable ABC trans
11	132.5	4.3	286	A72582	probable A/G-speci
12	132	4.3	1819	A71928	cag island protein
13	131.5	4.3	997	T43523	cut17 protein - fi
14	130	4.3	1927	G64585	cag pathogenicity
15	128.5	4.2	946	A96748	hypothetical prote
16	128	4.2	853	T51505	hypothetical prote
17	126	4.1	561	H86442	hypothetical prote
18	124.5	4.1	491	T50346	unknown protein [i
19	124	4.1	650	T33350	hypothetical prote
20	124	4.1	1702	T14050	hypothetical prote
21	122.5	4.0	669	S55024	protein kinase (EC
22	122	4.0	651	C86333	nebulin, skeletal
23	122	4.0	3924	S37431	hypothetical prote
24	120.5	3.9	911	S51441	ankyrin 2, neurona
25	120.5	3.9	1147	JN0599	hypothetical prote
26	120.5	3.9	1233	S56271	DNA-binding protei
27	120	3.9	1131	A49393	hypothetical prote
28	120	3.9	1440	T33813	activator 1 large
29	120	3.9	1658	S55101	hypothetical prote

30	120	3.9	2253	2	T30336	nuclear/mitotic ap
31	119.5	3.9	1740	2	T43773	hypothetical prote
32	119.5	3.9	3122	2	T17202	DNA-directed DNA p
33	119	3.9	646	2	F71620	hypothetical prote
34	119	3.9	797	2	H84919	hypothetical prote
35	119	3.9	991	2	H86168	hypothetical prote
36	119	3.9	1002	2	T30546	major surface glyco
37	118.5	3.9	891	2	H84614	hypothetical prote
38	118.5	3.9	1040	2	E71412	hypothetical prote
39	118	3.9	533	2	B84590	hypothetical prote
40	118	3.9	734	2	B42680	nucleolus-cytoplas
41	118	3.9	1200	2	A46194	neurofilament prot
42	118	3.9	1377	2	T51447	transcription regu
43	118	3.9	3329	2	T42205	breast cancer susc
44	118	3.9	3329	2	T30904	breast cancer tumo
45	117.5	3.8	957	2	T10633	hypothetical prote

ALIGNMENTS

RESULT 1
S57963
methyl CpG binding protein 2 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 05-Nov-1999
C:Accession: S57963
R:d'Esposito, M.; Quaderi, N.A.; Ciccodicola, A.; Bruni, P.; Esposito, T.; D'Urso, M.
submitted to the EMBL Data Library, July 1995
A:Description: Physical mapping and expression analysis of an X-linked gene encoding
A:Reference number: S57963
A:Accession: S57963
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-476 <DES>
A:Cross-references: EMBL:X89430; NID:g899295; PIDN:CAA61599.1; PID:g899296

Query Match 6.8%; Score 209; DB 2; Length 476;
Best Local Similarity 24.0%; Pred. No. 4.1e-06;
Matches 104; Conservative 53; Mismatches 149; Indels 128; Gaps 17;

QY	37	EDVAMELE	RVGEDEQMMTKRSECNPLIQEPTASAQFG-----ATAGTE	81
DB	13	KDKPLAFK	VKKVKDKKEKEKHEPVQPSAHSEPAEAKAETSEGGSGARLCEASAPK	72
QY	82	CRKSV-----	PGWERNVVKQRLFGKTAGRFVYFISPOGLKFRKSSILANYL	128
DB	73	QRRSILRD	RGPMDPTLPFGWTRKLLQKRSAGKYDVLINPQKAFRSKVELIAVF	132
QY	129	HKNGETS	LKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNSNNWNLRTSRCKKDV	188
DB	133	ERVGDTSL	DPNDFDFTV-TGRGSPSR-----REQKPKK--	165
QY	189	FMPPPSSS	SELQESRGL----SNFTSTHLLKDEGVDVDFNFRKVRKPKGVITLKGPIPKK	245
DB	166	---PKSPK	APAGTGRGRGKPGSGSTTRPKAATSEGUVVK--RVLEKSPGK--LLVMPPF-Q	217
QY	246	TKKGCKK	SCSGFVQSDS----KREYCNKADAESEPAQKSQLDRTVCISDAGACGETUS	301
DB	218	TSPGGKAE	GCGGATTSTQVMVVKRPGRKRAEADPOAIPKKR-----GRKPGSVVA	267
QY	302	VTSEENS	LKKKERSLSSCSNFCSEOKTSGIINKFCSAKDSEHNEKYEDTFLESEIGTK	361
DB	268	AAAAEAK	KAVKRESSIR-----SVQETVLPKK-----RKTRT-----VS	303
QY	362	VEVVERKE	HLHTDIL--KRGSEMDNNCSPTRDFTGEKIFQEDTIPRTOIERRRKTSLYFS	419
DB	304	IEVKVVK	PLLVSTLCEKSGKGLKTKCKSPGRK-----SKESPGR-----	344
QY	420	SKYNKEAL	SPRRK 433	
DB	345	---SSAS	ASSPPKKE 355	

A:Description: The SAC3 gene codes for a nuclear protein required for normal mitosis.

A:Reference number: S51323

Accession: S51323

A:Molecule type: DNA

A; Residues: 1-1301 <BAU>

A;Cross-references: EMBL:Z47805; NID:q634085; PID:q634086

R; Murphy, L.; Richards, C.; Harris, D.

submitted to the EMBL Data Library, July 1995

A:Reference number: S57971

A:Accession: S57983

A:Molecule type: DNA

A:Residues: 1-1301 <MUR>

A:Cross-references: EMBL:250046; NID:q899393; PID:q899406; MIPS:YDR159w

A: Experimental source: strain AB972

R: Bauer, A.: Koelling, R.

Yeast 12. 965-975. 1996

A:Title: Characterization of the SAC3 gene of *Saccharomyces cerevisiae*

A: Reference number: S71744; MIID: 97027306

A: Accession: S71744

A: Molecule type: DNA
Accession: 371744

A:Residues: 1-1301 <RAW>

A: Cross-references: EMBL: Z47805; NID: a634085; PID: CAA87767.1; PID: a634086

C:Genetics:

A-Gene: SGD-SAC3-LEP1

A/Selle: 300.0AC3, 4411
A-Cross-references: SGH-S0002566, MIPS-YDR159W

A: CROSS REFERENCES: 500.50002500, 510.5.1015300
A: Map position: 4R

R: Map position, 4K
C: Function.

A:Description: potential regulator of leucine permease gene expression

A; description: potential regulator of leucine permease expression

C;keyWords: nucleus; transmembrane protein
E:000-1015/Domains: transmembrane #status predicted <TMM>

Query Match	4.7%	Score 145;	DB 2;	Length 1301;
Best Local Similarity	21.9%	Pred. No. 0.22;		
Matches 116:	Conservative	61;	Mismatches 185;	Indels 168;
				Gaps 25;

[illegible]

RESULT 2

A41907 methyl-CpG-binding protein 2 - rat
N:Alternate names: chromosomal protein MeCP2
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 08-Oct-1999
C:Accession: A41907 #S41461
R:Lewis, J.D.; Meehan, R.R.; Henzel, W.J.; Maurer-Fogy, I.; Jeppesen, P.; Klein, F.; Bird
Cell 69, 905-914, 1992
A:Title: Purification, sequence, and cellular localization of a novel chromosomal protein
A:Reference number: A41907; MUID:92298389
A:Accession: A41907
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-492 <LEW>
A:Cross-references: GB:M94064; NID:g205361; PIDN:AAA1584.1; PID:g205362
R:Nan, X.; Meehan, R.R.; Bird, A.
Nucleic Acids Res. 21, 4886-4892, 1993
A:Title: Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2.
A:Reference number: S41461; MUID:94232813
A:Contents: annotation; methyl CpG-binding domain
C:Keywords: chromosomal protein; DNA binding
F:78-162/Domain: methyl-CpG-binding #status experimental <MCG>

Query Match	6.6%	Score	201;	DB	2;	Length	492;
Best Local Similarity	22.9%	Pred.	No. 1.4e-05;				
Matches	108;	Conservative	57;	Mismatches	150;	Gaps	18;

Qy	34	LRKE-	-----DVAMELERVGEDEEOMMIKRSECNPLLQEPITASAQFG-----	75
Db	8	LRKEKSEDDLOGLKPKLFFKKVKKDKKEDCKEHPLOPSAHHSAPRAEAGKETS	67	
Qy	76	-----ATAPTECRKSV-----	PCGWERVVVKORLFCKTAGREDVYFSP	113
Db	68	SGSAPVPEASASPKQRRIIRDGRPMYDDPTLP	EGWTRKLRKQKSGRSAGKYDVYILNP	127
Qy	114	QGLKFRKSSILANYLHKNGETSLKPEDFTFVLSKRGKISRYKDCSMAALTSHLONOSNN	173	
Db	128	QGFARSKVELIAYFEKVGDTSLDPNDFTV	TGRGSPSR-----	167
Qy	174	SNWNLRTRSKCKDVFMPPSSSSSLOBSRGLSNFTSTHLLKKEGVDVDFNFRKV	RKPK	232
Db	168	-----REOKPCK-----	PKSPKAPGTGRGRGPKGTGRPKAAASEGVQVKRLEKSP	217
Qy	233	GKVITLKGIPKTKKCKRSCSGFVSDS-----	KRESCVNKADAESPEVAOKSOLDRT	287
Db	218	GK--LLVKMPPQASPGG--KEGGGATTSQAQVWIKRGRKRAEADPQAIPKRR-----	268	
Qy	288	VCI SDAGACGTTLSVTSEENSVLVKKKRSISSGSNFCSEOKTSGIINKFCSAKDSEHNEK	347	
Db	269	-----GRKPGSVAAMAAEAKKVAKESSIR-----	SVQETVLPKIK-----RK	307
Qy	348	YEDTFLESEEIGTKVEVVERKEHLHTDIL--KRGSEMDNCCSPTRKDFTEKIFQEDTIP	405	
Db	308	TRET-----VSIEQEVVVKPLVLTGLKESGKLUKTCCKSPGRK-----	SKESP	351
Qy	406	RTQIERRKTSLYFSKYNKEALSPPRR-----	KAFKWKTPPRSP	444
Db	352	KGR-----SSSASSPPKKEHHHHHHHAESPAPMPLLLPPPPP	388	

RESULT 3

SAC3 protein - yeast (*Saccharomyces cerevisiae*)
S51323

N:Alternate names: protein YD8358.13; protein YDR159w
C:Species: *Saccharomyces cerevisiae*
C:Date: 23-Feb-1995 #sequence_revision 12-May-1995 #text_change 21-Jul-2000
C:Accession: S51323; S57983; S71744
R:Bauer, A.; Koelling, R.
Submitted to the EMBL Data Library, January 1995

RESULTS

G02334
breast cancer tumor suppressor BRCA2 - human
N;Alternate names: breast cancer susceptibility protein BRCA2
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 22-Jun-1999
C;Accession: G02334; S68501
R;Tavtigian, S.V.; Rommens, J.M.; Couch, F.J.; Neuhausen, S.; Bell, R.; Berry, S.; Bozdeger, M.; Snyder, S.; Stringfellow, M.; Stroup, C.; Swedlund, B.; Teng, D.; Thomas, A.; Titchard, K.; Vogelstein, B.; Vogelstein, N.
Submitted to the EMBL Data Library, December 1995
A;Reference number: H01078
A;Accession: G02334
A;Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A;Residues: 1-3418 <TA>
A;Cross-references: EMBL:U43746; NID:glll61383; PID:NAB07223.1; PID:glll61384
I;Haworth, P.A.; Connor, F.; Arason, A.; Gudmundsson, J.; Ficenc, D.; Kellsell, D.; Ford, D.; Taylor, N.; Narod, S.; Lenoir, G.; Egilsson, V.; Barkadottir, R.B.; Easton, D.F.; Bentley, D.R.
Nature 378, 789-792, 1995
A;Authors: Futreal, P.A.; Ashworth, A.; Stratton, M.R.
Title: Identification of the breast cancer susceptibility gene BRCA2.
A;Reference number: S68501; MOID:96112016
A;Accession: S68501
A:Molecule type: mRNA
A;Residues: 282-371, 'N', 373-598, 'S', 600-1108, 'EQ', '1111-1119, 'D', '1121-2321, 'V', 2323-2386; I;Gentiles:
A;Gene: GDB:BRCA2
A;Cross-references: GDB:387848; OMIM:600185
A;Map position: l3ql2.3-l3ql2.3
C;Superfamily: breast cancer tumor suppressor BRCA2
C;Keywords: polymorphism; tumor suppressor

C:\Users\Public\Desktop\BIOINFORMATICS\GenBank Files

[illegible]

RESULTS

RESOL
S27833

rhoptry-associated protein 1 precursor - malaria parasite (Plasmodium falciparum)
N:Alternate names: protective antigen
C:Species: Plasmodium falciparum
C:date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jun-2000
C:Accession: A45514; S27833
R:Ridley, R.G.; Takacs, B.; Lahm, H.W.; Delves, C.J.; Goman, M.; Certa, U.; Matile, H
Mol. Biochem. Parasitol. 41, 125-134, 1990
A:title: Characterisation and sequence of a protective rhoptry antigen from Plasmodium
A:Reference number: A45514; MUID:90348711
A:Accession: A45514
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-782 <R12>
A:Cross-references: GB:M32853; NID:gl160656; PID:gl160657
C:Superfamily: Plasmodium falciparum rhoptry-associated protein 1

Query Match	4.6%;	Score 141.5;	DB 2;	Length 782;
Best Local Similarity	21.1%;	Pred. No. 0.18;		
Matches 110; Conservative 71;	Mismatches 176;	Indels 165;	Gaps	

QY	123	SLANYLHKNGET---	SLKPEDDF-----	TVLSRKGISRYKDCSMAALTSHLONQS-	171
	:	:::	:::	:::	:
Db	18	NVADGGINVGNDNNYGKTIINDDNFDDYNWTPINKKEFLNSYED--EFSSSESFLENKS	S	75	
QY	172	-NASNWNLRTRSRCKKDVFMPPPSSSELQESRGLSNFTSTHLLLKLEDEGVDDVNFRKVRK	230		
	:	:::	:::	:::	:
Db	76	VDDGNINLTDTSTSNK-----	SSKHGRSVRSASAAAILEEDSKDMEP----	122	
QY	231	PKGVVTILKGIPTKYTKGCRKSCSGFVSODSKRESVCNKADAESPVAQKSQLDRTVCI	290		
	:	:::	:::	:::	:
Db	123	-KASPVSVKITSPSGTQTGLKSS--PSSTKSSPSN-----	V	158	
QY	291	SDAGACGETLSVTSSENSLVKKERSLSSGSNFCSEOKTSGIINKFCSAKDSEHNEKEYE	350		
	:	:::	:::	:::	:
Db	159	KSPHGES---NSSSESTTKSKRSAS-----	VAGIV-----	GADGEAPPAKN	200
QY	351	TFLESEEI-GTKVEVVERKEHLJT---	DILKRG-----	EMDNCSPTRKDFTG	395
	:	:::	:::	:::	:
Db	201	TLTPLPLEELYPNVNLFNKYSLSNMMEENINLLKNEGDLVAQKEEPYDENMEKAKDKKK	260		
QY	396	--EKI-----	FOEDITPTQTERKTSLYSFYSSKNKALSPPRRKAFKMTTPPRS	443	
	:	:::	:::	:::	:
Db	261	ALEKIGKQSDPEPFMFSENKFLVENQKERNVAGSFRRFSK-LNP----	EKK-----	307	
QY	444	PFNLVOETFPHDPWKLLIATIFLNRTSGKMAIPVLWKFLKPYSA-----	488		
	:	:::	:::	:::	:
Db	308	-DEVIEKT---EVSKRTFGSIGFNLTDKBAKVILGVGATGYOETPMLYNCPNNSNLFDTPI	363		
QY	489	-----	EVARTADWRDVSELLPLGLYLDIRAK-TIVFKSDEVULTKWYPIEL	534	
	:	:::	:::	:::	:
Db	364	ESLQGRIIDIKKRESMLSTTFEOQKCLNMGMVDLELNDTOCKGT-----	410		
QY	535	HGIGKYGNDOSYRI-F-CVNEKWQHVP-----	EDHKLKNYH	567	
	:	:::	:::	:::	:
Db	411	-CIGSGSEGHILLRIEFDENLFKHPHDIDYLTLDADGYKLOKNH	451		

RESULT 6
nucleolar protein gar2 - fission yeast (Schizosaccharomyces pombe)
S55785
N:Alternate names: probable RNA-binding protein RBD18
C:Species: Schizosaccharomyces pombe
C:Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 17-Nov-2000
C:Accession: S55785; S68087; T37634; S52318
R:Gulli, M.P.; Girard, J.P.; Zabetakis, D.; Lapeyre, B.; Melese, T.; Caizergues-Ferre
Nucleic Acids Res. 23, 1912-1918, 1995
A:Title: gar2 is a nucleolar protein from Schizosaccharomyces pombe required for 18S
A:Reference number: S55785; MUID:95319932
A:Accession: S55785
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-500 <GUL>

RESULT 5
S27833


```

Db 159 KDDG-----SEALDDGI-----ESSEKDKKKKKKSENDADAPKKDRKTRKKBEKARKLA 208
Qy 373 TDILKRGSEMDNNCSPTRKD-----FTGEKIFQEDTIPRTQTERRKTSL 416
Db 209 SMLESENKDNDANAAPLNKTDFAKDLPSGRLLIFAYASGQKVPADGSPADGI-----261
Qy 417 YFSSKYNKEALSPRRPKAFKKTPPRSPFNLVQETLFHDPW-KLLIATIFLNRTSGK 472
Db 262 ---TVTGNLLSPNSR---LQVEKLSVSAWGKLLIKDSELNLNGR 302

RESULT 11
A72682
probable A/G-specific adenine glycosylase APE0875 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: A72682
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kikuchi, T.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix strain K1
A:Reference number: A72450; MUID:99310339
A:Accession: A72682
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-286 <KAW>
A:Cross-references: DDBJ:AP000060; NID:G5104188; PIDN:BAA79857.1; PID:d1043643; PID:g5104188
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0875

```

	Query Match	Best Local Similarity	4.3%;	Score 132.5;	DB 2;	Length 286;
	Matches	28;	Conservative	26;	Mismatches	31;
					Indels	9;
					Gaps	3;
QY	455	DPWKLIIATITLARTSCMAIPVLWKRFLKYPKSAEAVRTADRDVSELLPLGLYDLRAK	514			
DB	98	DPAWILVAAPFLRKTTTARQVVRVVEEFLRYPNPKALASAREDEVRELIIRPLGIEHORAK	157			
QY	515	TIYKFSDEYLTQW--KYPI-----ELHGIGKY	540			
DB	158	HLIETLA-KHIEARYGGRIKPCSKKELPLGVGY	190			

RESULT 12
A71928
cag island protein - *Helicobacter pylori* (strain J99)
C:Species: *Helicobacter pylori*
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: A71928
R:Alm, R.A.; Ling, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557
A:Accession: A71928
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1819 <ARN>
A:Cross-references: GB:AE001481; GB:AE001439; NID:g4155005; PIDN:AAD06047.1; PID:g415500
A:Experimental source: strain J99
C:Genetics:
A:Gene: orf13/14

```
Query Match          4.3%; Score 132; DB 2; Length 1819;
Best Local Similarity 18.2%; Pred. No. 2.4;
Matches      87; Conservative   83; Mismatches 183; Indels    124; Gaps     19;
```

Db 942 LENOALDC LKNAKTEAEKKRCVKDLPKDLQKKVLAKESVRYVYDCVSKAKNAEAKKECEK 1001
 Qy 50 -----EQOMMIKRSECNPLLPQEP IASAQFGATAGTECKRSVPCGWGVRVVKORLFGKTAGR 105
 Db 1002 LLTPEARKLLEAEKESVKAYKDCVSRAR--NEKEKOECEKLLTTP EARKLLEOEVKKSVKAY 1060
 Qy 106 FDVY-----PISQGLKFRSKSSILANYLHKNETSLK-----PEPDF 143
 Db 1061 LDCVSRAFNEKEKOECEKLLTTP EARKLLENQALD--CLKNAKTEAEKKRCVKDLPKDLQK 1118
 Qy 144 TVLSKRGIKSRVKKCSMAALTSHLQNSNNSNNILRTSKCKDKDVFMPSPSSSELQESRG 203
 Db 1119 KVLAKESVKA-YLDC-----VSRARNE-----KEKKECEK--LTTPEARKLLBESK- 1161
 Qy 204 LSNETSTHL--LLKEBEGVDDVFNFRVKRPKGVKVTILKGIPIKTKKC----- 250
 Db 1162 --KSVRAYLDCVSKAKNAEAKKECEKLLTTP EARKLLEAEKESVKAYKDCVSRARNEKEKQ 1219
 Qy 251 -----BKSGGFVQSDSKSEVCNRADAES--EPA-----QRSQL 284
 Db 1220 ECBKLLTTP EARKLLEOEVKKSVKAYLDCVSRARNEKEKOECEKLLTTP EARKLEKQROK 1279
 Qy 285 DRVV--CI-----SDAGAGEVLSVTSEENSLVKKKERSLSSGNSFCSEQKTSGLINKFC 337
 Db 1280 DKAILKCKLNADPNRAAIMKCLDGLSDEKL-----KYLOFAREKAVLDCLK 1327
 Qy 338 SAKDSEHNEKYEDTFLE--SEETGTVEVYVERKEHLTDLIKRGSE--MDNNCSPTRKD 392
 Db 1328 TARTDEEKRKCNLYSDLIQETONKKAQNQKLSKTELHQASECLNDLDDPTDQE 1384
 RESULT 13
 T43523
 cut17 protein - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 21-Jan-2000 #sequence revision 21-Jan-2000 #text_change 02-Jun-2000
 C:Accession: T43523; T41649; T41700
 R:Morishita, J.; Matsusaka, N.; Yanagida, M.
 submitted to the EMBL Data Library, August 1999
 A:Description: Fission yeast cut17 is required for chromosome segregation.
 A:Reference number: Z22536
 A:Accession: T43523
 A:Status: translated from GB/EMBL/DBDJ
 A:Molecule type: mRNA
 A:Residues: 1-997 <MOR>
 A:Cross-references: EMBL:AB031034; PIDN:BAA83415.1
 R:Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, August 1998
 A:Reference number: Z22007
 A:Accession: T41649
 A:Status: translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-997 <HAR>
 A:Cross-references: EMBL:AL031323; PIDN:CAA20434.1; GSPDB:GN000068; SPDB:SPC
 A:Experimental source: strain 972h; cosmid c962
 R:Wedler, H.; Buesterhoef, A.; McBoudail, R.C.; Rajandream, M.A.; Barrell,
 submitted to the EMBL Data Library, October 1999
 A:Reference number: Z22010
 A:Accession: T41700
 A:Status: translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 932-997 <WED>
 A:Cross-references: EMBL:AL121859; PIDN:CAB58376.1; GSPDB:GN000068; SPDB:SPC
 C:Genetics:
 A:Gene: cut17; SPCC962.02c; SPDB:SFCP31B10.10c
 A:Map position: 3L
 A:Insertions: 43/3

Query Match 4.3%; Score 131.5; DB 2; Length 997;
 Best Local Similarity 20.1%; Pred. NO. 1.1;
 Matches 129; Conservative 119; Mismatches 241; Indels 153; Gaps 33;

Db 531 SRSMKKYQSEDEDEISEQKPRAKDKAASGOQSVGSISSRSQARKLQPNKMMSSITPR 590
Qy 227 ---KVRKPKGKVTILKGIPIKTKGCRKSCGFVQSDSKRESVCNKADAESEPVAKSQ 283
Db 591 SAASVPRPSGKVS-----NTSSGRRS-----DKSLAQSVNPFSELIKENTKPSL 636
Qy 284 LDRVCISDAGACGETLSVTSEENSLVKKE--RSLSSG-----SNFCSEQKTSIIN 334
Db 637 AVKTTMRSOVKSSGRTKNI--KEDTLQRPRLKSSSGNIDFTELSTLCSDDM---MVS 691
Qy 335 KFCSAKDSE--HNEKYEDTFLESEI--GKVEVVERKEHLHTDILKRGSEM----- 382
Db 692 LRVDSDISETLRNEEYDEPAEPEEVLNAREEEVEELETILVPEDGNPMLSEAYEKVD 751
Qy 383 ---DNNGSPTRKDFTEKIFQEDTIPRT 407
Db 752 HSGENC-----FLPATVPTT 768

Search completed: January 15, 2002, 13:52:37
Job time: 406 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: January 15, 2002, 13:51:13 ; Search time 57.72 Seconds
(without alignments)
123.589 Million cell updates/sec

Title: US-09-724-296-37

Perfect score: 1662

Sequence: 1 MIFRFGVSNAMSLWDASPA.....ELSSIRGVKRRIGGALQWKS 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2.6/ptodata/2/iaa/5A_COMB.pep:.*
2: /cgn2.6/ptodata/2/iaa/5B_COMB.pep:.*
3: /cgn2.6/ptodata/2/iaa/6A_COMB.pep:.*
4: /cgn2.6/ptodata/2/iaa/6B_COMB.pep:.*
5: /cgn2.6/ptodata/2/iaa/PCTUS_COMB.pep:.*
6: /cgn2.6/ptodata/2/iaa/backfiles1.pep:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	93.5	5.6	1041	3	US-08-898-978-2
2	91.5	5.5	522	1	US-08-625-322-2
3	84.5	5.1	1073	4	US-09-541-782-6
4	84.5	5.1	1255	3	US-08-947-823-3
5	83.5	5.0	438	3	US-08-886-886-11
6	82.5	5.0	500	1	US-08-704-398-2
7	82.5	5.0	500	5	PCT-US95-05966-2
8	82.5	5.0	2710	1	US-08-480-604A-6
9	82.5	5.0	2710	2	US-08-405-496A-6
10	82.5	5.0	2710	4	US-08-515-136-6
11	82	4.9	323	6	5185254-4
12	81.5	4.9	892	1	US-07-977-434-12
13	81.5	4.9	892	1	US-08-458-819-12
14	81.5	4.9	892	5	PCT-US91-07035-12
15	79.5	4.8	859	1	US-07-854-596B-35
16	79	4.8	376	4	US-09-041-718-2
17	79	4.8	3224	2	US-08-705-660-34
18	79	4.8	3224	3	US-08-989-045-34
19	78.5	4.7	410	4	US-09-258-754-450
20	78.5	4.7	615	1	US-08-484-105-6
21	78.5	4.7	615	1	US-08-484-106-6
22	77.5	4.7	483	3	US-09-027-166-7
23	77	4.6	704	1	US-08-188-582-18
24	77	4.6	704	1	US-08-646-715-18
25	77	4.6	2293	4	US-09-368-590-2
26	76.5	4.6	772	1	US-08-524-757-12
27	75.5	4.5	396	6	5349058-2

28	75	4.5	704	1	US-08-188-582-5
29	75	4.5	704	1	US-08-646-715-5
30	75	4.5	1257	3	US-08-947-823-5
31	74.5	4.5	369	1	US-08-844-055-2
32	74.5	4.5	369	3	US-09-006-849-2
33	74.5	4.5	381	1	US-07-936-163-3
34	74.5	4.5	429	1	US-08-906-744A-2
35	74.5	4.5	429	3	US-09-093-134-2
36	74.5	4.5	541	3	US-08-890-615-2
37	74.5	4.5	976	4	US-09-104-324B-4
38	74.5	4.5	2522	4	US-09-251-645-13
39	74	4.5	263	4	US-09-159-106-2
40	74	4.5	303	4	US-09-159-106-13
41	74	4.5	435	4	US-09-159-106-11
42	74	4.5	436	4	US-09-150-213-4
43	74	4.5	2496	4	US-09-125-028-2
44	74	4.5	2958	4	US-08-894-344C-2
45	73.5	4.4	264	2	US-08-402-804-4

ALIGNMENTS

RESULT 1

US-08-898-978-2
; Sequence 2, Application US/08898978
; Patent No. 6001602
; GENERAL INFORMATION:
; APPLICANT: Reichard, Raymond W.
; APPLICANT: Brown, James R.
; APPLICANT: Lawlor, Elizabeth J.
; TITLE OF INVENTION: No. 6001602el files
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/898,978
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: GM10051
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-898-978-2

Query Match 5.6%; Score 93.5; DB 3; Length 1041;
Best Local Similarity 19.7%; Pred. No. 0.39;
Matches 57; Conservative 42; Mismatches 104; Indels 87; Gaps 13;


```

RESULT 6
US-08-704-398-2
; Sequence 2, Application US/08704398
; Patent No. 5679525
; GENERAL INFORMATION:
; APPLICANT: Peterson, Michael G
; APPLICANT: Henkel, Thomas
; TITLE OF INVENTION: EPSTEIN-BARR VIRUS TRANSCRIPTION FACTOR
; TITLE OF INVENTION: BINDING ASSAY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOUBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,398
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/246,977
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-592333/RAO
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-704-398-2

```

```

Query Match 5.0%; Score 82.5; DB 1; Length 500;
Best Local Similarity 21.6%; Pred. No. 1.9;
Matches 51; Conservative 32; Mismatches 68; Indels 85; Gaps 12;

QY 120 PRESVTKNAVTDMAHYRMLEAMGADRSVINIHGGA---YGN----- 160
Db 27 PKRLITREARN-----YKRG--DQVLLIHAKVAQKSYGNEKRFPCPPCVYLMGS 78

QY 161 -----KDTATAQFHQ-----NIKQLPOEIKERMTLENDKTYTTEETLQVCEQED 205
Db 79 GWKKKKEQMERDCSEOSQPCAFIGNSDQEQ---LNLEKKNYCTAKTLYISDSK 135

QY 206 -VPFVDFHFFYANPDHADLNVALPRMIKTWIGLQPKVHLSSP-KSQAIRSHADYV 263
Db 136 RKHIFSVKMFYGNSD---DIGVFLSKRIK-----ISKPSKKQSLK----- 175

QY 264 DANFLLERFRQWNTIDPMIEAKQKDKALLRLMDLSSIRGVKRIKG---GALOW 315
Db 176 -----NADLCIASGTKVALFNRLRSQTVSTRYLHVEGNGFIASSQQW 217

```

```

RESULT 7
PCT-US95-05966-2
; Sequence 2, Application PC/TUS9505966
; GENERAL INFORMATION:
; APPLICANT: TULARIK, INC.
; TITLE OF INVENTION: EPSTEIN-BARR VIRUS TRANSCRIPTION FACTOR

```

```

; TITLE OF INVENTION: BINDING ASSAY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOUBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05966
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/246,977
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: FP-59233-PC/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-05966-2

```

```

Query Match 5.0%; Score 82.5; DB 5; Length 500;
Best Local Similarity 21.6%; Pred. No. 1.9;
Matches 51; Conservative 32; Mismatches 68; Indels 85; Gaps 12;

QY 120 PRESVTKNAVTDMAHYRMLEAMGADRSVINIHGGA---YGN----- 160
Db 27 PKRLITREARN-----YKRG--DQVLLIHAKVAQKSYGNEKRFPCPPCVYLMGS 78

QY 161 -----KDTATAQFHQ-----NIKQLPOEIKERMTLENDKTYTTEETLQVCEQED 205
Db 79 GWKKKKEQMERDCSEOSQPCAFIGNSDQEQ---LNLEKKNYCTAKTLYISDSK 135

QY 206 -VPFVDFHFFYANPDHADLNVALPRMIKTWIGLQPKVHLSSP-KSQAIRSHADYV 263
Db 136 RKHIFSVKMFYGNSD---DIGVFLSKRIK-----ISKPSKKQSLK----- 175

QY 264 DANFLLERFRQWNTIDPMIEAKQKDKALLRLMDLSSIRGVKRIKG---CALOW 315
Db 176 -----NADLCIASGTKVALFNRLRSQTVSTRYLHVEGNGFIASSQQW 217

```

```

RESULT 8
US-08-480-604A-6
; Sequence 6, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:

```

ADDRESSEE: MEDLEN & CARROLL, LLP
 STREET: 220 MONTGOMERY STREET, SUITE 2200
 CITY: SAN FRANCISCO
 STATE: CALIFORNIA
 COUNTRY: UNITED STATES OF AMERICA
 ZIP: 94104

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/480,604A
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 424

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/422,711
 FILING DATE: 14-APR-1995

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/405,496
 FILING DATE: 16-MAR-1995

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/329,154
 FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/161,907
 FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/985,321
 FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/429,791
 FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:
 NAME: INGOLIA, DIANE E.
 REGISTRATION NUMBER: 40,027
 REFERENCE/DOCKET NUMBER: OPHD-01763

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2710 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-08-480-604A-6

Query Match 5.0%; Score 82.5; DB 1; Length 2710;
 Best Local Similarity 18.2%; Pred. No. 29;
 Matches 61; Conservative 54; Mismatches 116; Indels 105; Gaps 17;

QY 29 SKLSKTERKEALLVTYKANLNTMRTLHYI-IGHGIPLYRFSSSIVPLATHDPVMDVFT 87
 DB 72 SNLKKDILKEVIL-INKNSPTSPEVKNLHFVWIGGEVS-----DIAL 111

QY 88 PFOKEFREIGELVKTHOLRTSFHPNQFTLTSPK---ESVTKNV-----TDM 132
 DB 112 EYIKQADIN---AENIKLWYDSEAFVNTLTKKAIVESSTTEALQLLEEIONPQFDNM 168

QY 133 AYHYRMLEANGIADR-----SVINIHIGGAYGNKDTATAQFHON 171
 DB 169 KFYKKRMEF--IYDROKRFINYKSKQINKPTVPTIDDIKLSHLVSEYNRDETVLESYRTN 226

QY 172 IKOLPOEIKRWTLN-NDKTYTTEETLOVCEOE-----DVPEVDFDHFH-YY 216
 DB 227 SL---RKINSNHGIDIRANSLETEQELLNYSOELLNRGNLAAASDIVRLLALKNFGVY 283

QY 217 AN----PDHADLNLVLPW----IKTWERIGLOPKVHLSSPKSCAIRSHADYVDANFL 268
 DB 284 LDVDMPLPGHISDLFTKTSRSPSSIGLDRWEMIKL-----EAINKYKKYIN-NYT 330

QY 269 LERF-----ROWGTNIDPFMLEAKQDKKALLRLMDELS 300
 DB 331 SENFDKLDQDLKDNPKLJIESKSEKSEIFSKLENLN 366

RESULT 9
 US-08-405-496A-6
 ; Sequence 6, Application US/08405496A
 ; Patent No. 5919665
 ; GENERAL INFORMATION:
 ; APPLICANT: WILLIAMS, JAMES A.
 ; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
 ; TITLE OF INVENTION: NEUROTOXIN
 ; NUMBER OF SEQUENCES: 30
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MEDLEN & CARROLL, LLP
 ; STREET: 220 MONTGOMERY STREET, SUITE 2200
 ; CITY: SAN FRANCISCO
 ; STATE: CALIFORNIA
 ; COUNTRY: USA
 ; ZIP: 94104

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/405,496A
 FILING DATE: 16-MAR-1995
 CLASSIFICATION: 424

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/329,154
 FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/161,907
 FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/985,321
 FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/429,791
 FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:
 NAME: INGOLIA, DIANE E.
 REGISTRATION NUMBER: 40,027
 REFERENCE/DOCKET NUMBER: OPHD-01308

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2710 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-08-405-496A-6

Query Match 5.0%; Score 82.5; DB 2; Length 2710;
 Best Local Similarity 18.2%; Pred. No. 29;
 Matches 61; Conservative 54; Mismatches 116; Indels 105; Gaps 17;

QY 29 SKLSKTERKEALLVTYKANLNTMRTLHYI-IGHGIPLYRFSSSIVPLATHDPVMDVFT 87
 DB 72 SNLKKDILKEVIL-INKNSPTSPEVKNLHFVWIGGEVS-----DIAL 111

QY 88 PFOKEFREIGELVKTHOLRTSFHPNQFTLTSPK---ESVTKNV-----TDM 132
 DB 112 EYIKQADIN---AENIKLWYDSEAFVNTLTKKAIVESSTTEALQLLEEIONPQFDNM 168

QY 133 AYHYRMLEANGIADR-----SVINIHIGGAYGNKDTATAQFHON 171
 DB 169 KFYKKRMEF--IYDROKRFINYKSKQINKPTVPTIDDIKLSHLVSEYNRDETVLESYRTN 226


```

; FILING DATE: 22-AUG-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 746,121
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US90/07641
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 585,471
; FILING DATE: 20-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 455,611
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 609,157
; FILING DATE: 02-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 557,517
; FILING DATE: 24-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Luann Cserr
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: Case No. 5795762 8753
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2972
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 892 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-458-819-12

```

Query Match 4.9%; Score 81.5; DB 1; Length 892;

Best Local Similarity 17.1%; Pred. No. 6.2; Mismatches 116; Indels 109; Gaps 11;

```

; Matches 58; Conservative 57; Mismatches 116; Indels 109; Gaps 11;
Qy 31 LSKTERKEALLTVTKANLRNMTLHYIIGHGIPLYRFSSSIPLATHPDVMMDFVTPFQ 90
Db 14 LNPDEKRNLEELSLKLYGKMFDELNVNVPFLGDFSYVPLERAVEYSCEDADVTY 473
Qy 91 KEFREIGELVTKHOLVTSFHPNQFTL-----116
Db 474 RIFKRLGRKIYENEMKLFYFIEMPLDIVSEMELNGVYFDEYKLSKKYQEKMDGK 533
Qy 117 ---FTSPKESVTKNAVDMAYHYRMLEAMGIADRSVINIHHGAYGNKDTATQAFHONIK 173
Db 534 EKVEIAGETENLSSQVA--YILFEKLNAP-----YKKTATGKFSTNAE 578
Qy 174 QLPQEIKE-----RMTLENDKTYTTEETLVQCEQEDVPFVFD-----FHH----- 214
Db 579 VLEELSKHEHIAKLLE-----YKYOGLKSTYIDSIPLSINRKTNRVHTFHTGTGTSTG 633
Qy 215 --FYANP-----DDHADLNVALPRMIKTW-----ERICLOPKVHLSPPKSEQAI 256
Db 634 RLSSNPMLQNLPTRESEKRIKAVRPQDDMWILGADYSQIELRVLAHVS--KDELNL 691
Qy 257 RSHADYVDAN-----FLLERFRWGTNIDFMI 283
Db 692 KAFKEDLDIHTAAKIFGVSEMFVDSQMRVRGKVMVFAI 731

```

RESULT 14

PCT-US91-07035-12

; Sequence 12, Application PC/TUS9107035

; GENERAL INFORMATION:

; APPLICANT: Gelfand, David H.

; APPLICANT: Abramson, Richard D.

; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF

; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES

; NUMBER OF SEQUENCES: 38

; CORRESPONDENCE ADDRESS:

```

; ADDRESS: Cetus Corporation
; STREET: 1400 Fifty-third Street
; CITY: Emeryville
; STATE: California
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/07035
; FILING DATE: 19910930
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,490
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,466
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,213
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 523,394
; FILING DATE: 15-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 143,441
; FILING DATE: 12-JAN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 063,509
; FILING DATE: 17-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,241
; FILING DATE: 22-AUG-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 746,121
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US90/07641
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 585,471
; FILING DATE: 20-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 455,611
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 609,157
; FILING DATE: 02-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 557,517
; FILING DATE: 24-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Sias Ph.D. Stacey R.
; REGISTRATION NUMBER: 32,630
; REFERENCE/DOCKET NUMBER: Case No. 2580
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-420-3300
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 892 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US91-07035-12

```

Query Match 4.9%; Score 81.5; DB 5; Length 892;

Best Local Similarity 17.1%; Pred. No. 6.2; Mismatches 116; Indels 109; Gaps 11;

Matches 58; Conservative 57; Mismatches 116; Indels 109; Gaps 11;

Qy 31 LSKTERKEALLTVTKANLRNMTLHYIIGHGIPLYRFSSSIPLATHPDVMMDFVTPFQ 90

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:50:07 ; Search time 120.01 Seconds
(without alignments)
195.661 Million cell updates/sec

Title: US-09-724-296-37
Perfect score: 1662
Sequence: 1 MIFRFGFVSNAWLDASPA.....ELSSIRGVKRIIGGALQWKS 317

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101.*

- 1: /SIDSB/gcgdata/geneseq/geneseq/AA1980.DAT.*
- 2: /SIDSB/gcgdata/geneseq/geneseq/AA1981.DAT.*
- 3: /SIDSB/gcgdata/geneseq/geneseq/AA1982.DAT.*
- 4: /SIDSB/gcgdata/geneseq/geneseq/AA1983.DAT.*
- 5: /SIDSB/gcgdata/geneseq/geneseq/AA1984.DAT.*
- 6: /SIDSB/gcgdata/geneseq/geneseq/AA1985.DAT.*
- 7: /SIDSB/gcgdata/geneseq/geneseq/AA1986.DAT.*
- 8: /SIDSB/gcgdata/geneseq/geneseq/AA1987.DAT.*
- 9: /SIDSB/gcgdata/geneseq/geneseq/AA1988.DAT.*
- 10: /SIDSB/gcgdata/geneseq/geneseq/AA1989.DAT.*
- 11: /SIDSB/gcgdata/geneseq/geneseq/AA1990.DAT.*
- 12: /SIDSB/gcgdata/geneseq/geneseq/AA1991.DAT.*
- 13: /SIDSB/gcgdata/geneseq/geneseq/AA1992.DAT.*
- 14: /SIDSB/gcgdata/geneseq/geneseq/AA1993.DAT.*
- 15: /SIDSB/gcgdata/geneseq/geneseq/AA1994.DAT.*
- 16: /SIDSB/gcgdata/geneseq/geneseq/AA1995.DAT.*
- 17: /SIDSB/gcgdata/geneseq/geneseq/AA1996.DAT.*
- 18: /SIDSB/gcgdata/geneseq/geneseq/AA1997.DAT.*
- 19: /SIDSB/gcgdata/geneseq/geneseq/AA1998.DAT.*
- 20: /SIDSB/gcgdata/geneseq/geneseq/AA1999.DAT.*
- 21: /SIDSB/gcgdata/geneseq/geneseq/AA2000.DAT.*
- 22: /SIDSB/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	359.5	21.6	371	21 AAY44499	S. pombe delta228-
2	359.5	21.6	600	21 AAY44500	GST signal peptide
3	359.5	21.6	828	21 AAY44498	GST signal peptide
4	345	20.8	656	17 AA001618	Neurospora crassa
5	313.5	18.9	626	21 AAY44502	N. crassa delta228
6	313.5	18.9	626	21 AAY44503	B. subtilis delta22
7	292.5	17.6	294	21 AAY44505	D. radiodurans del
8	103	6.2	964	18 AAW32619	Cyclic-isomaltolig
9	93.5	5.6	1040	20 AAW97216	An isoleucyl trna
10	93.5	5.6	1041	21 AAY56782	C. trachomatis ile
11	92.5	5.6	234	21 AAB03759	Human H37 amino ac

12	92.5	5.6	674	21 AAB42482	Human ORFX ORF2246
13	92.5	5.6	674	21 AAB03758	Human H37 amino ac
14	91.5	5.5	522	19 AAY77094	Sorting nexin 1.
15	91	5.5	558	21 AAY77979	A. thaliana enviro
16	91	5.5	1284	9 AAP81187	Sequence encoded b
17	90.5	5.4	652	19 AAW73029	Helicobacter pylor
18	89	5.4	261	18 AAW55332	H. pylori ORF hp3e
19	88.5	5.3	537	21 AAG14498	Arabidopsis thalia
20	88.5	5.3	549	21 AAG14497	Arabidopsis thalia
21	88	5.3	289	22 AAG81570	S. epidermidis ope
22	88	5.3	645	19 AAW73030	Helicobacter pylor
23	87.5	5.3	548	20 AAY37206	Protein involved i
24	87	5.2	1521	21 AAG39235	Arabidopsis thalia
25	87	5.2	1528	21 AAG39234	Arabidopsis thalia
26	87	5.2	1562	21 AAG39233	Arabidopsis thalia
27	87	5.2	1703	21 AAG36714	Arabidopsis thalia
28	87	5.2	1710	21 AAG36713	Arabidopsis thalia
29	87	5.2	1744	21 AAG36712	Arabidopsis thalia
30	85	5.1	398	22 AAG81475	S. epidermidis ope
31	85	5.1	398	22 AAG82202	S. epidermidis ope
32	84.5	5.1	249	18 AAW55445	H. pylori ORF Olae
33	84.5	5.1	249	20 AAY17212	H. pylori outer me
34	84.5	5.1	458	21 AAY97013	S. cerevisiae esse
35	84.5	5.1	1255	19 AAW55974	Tomato pest resist
36	83.5	5.0	438	22 AAG81537	S. epidermidis ope
37	83.5	5.0	1254	11 AAR07503	Meroprotease apical-e
38	83.5	5.0	1254	18 AAW24575	Meroprotease apical-e
39	83	5.0	806	22 AAB88467	Human membrane or
40	83	5.0	1619	22 AAU00970	Drosophila melanog
41	83	5.0	1619	22 AAU00986	Drosophila melanog
42	83	5.0	1619	22 AAU00987	Drosophila melanog
43	83	5.0	1619	22 AAU00988	Drosophila melanog
44	83	5.0	1704	22 AAU00969	Drosophila melanog
45	83	5.0	1704	22 AAU00983	Drosophila melanog

ALIGNMENTS

RESULT 1

- AAAY44499
- ID AAY44499 standard; Protein; 371 AA.
- AC AAY444499;
- XX
- DT 27-MAR-2000 (first entry)
- XX
- DE S. pombe delta228-UV damage endonuclease.
- XX
- KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
- KW glutathione-S-transferase signal peptide; uvel+ gene product;
- KW UV irradiation; DNA damage; UV radiation damage; photocproduct;
- KW abasic site; apatinum diaduct; mismatched nucleotide pairing;
- KW nucleotide alkylation; skin cancer.
- XX
- OS Schizosaccharomyces pombe.
- XX
- PN WO9963828-Al.
- PD 16-DEC-1999.
- XX
- PF 08-JUN-1999; 99WO-US12910.
- XX
- PR 08-JUN-1998; 98US-0088521.
- PR 18-MAY-1999; 99US-0134752.
- XX
- PA (UYEM-) UNIV EMORY.
- XX
- PI Doetsch PW, Kaur B, Avery AM;
- XX
- DR WPI; 2000-116417/10.
- DR N-PSDB; AAZ29859.
- XX

PT A new truncated ultraviolet damage endonuclease for treatment of skin
 XX cancers -
 PS Claim 13; Page 53; 133pp; English.
 XX
 CC The present sequence is delta228-UV damage endonuclease. Delta228-UVDE is
 CC a 288 amino acid deletion of the N-terminal of the S. pombe uvel+ gene
 CC product. This is expressed in frame with a GST leader sequence to
 CC generate a fusion protein. This provides stable endonuclease fragments
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, aplatium diaduct, an intercalated
 CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 XX Sequence 371 AA;
 SQ
 Query Match 21.6%; Score 359.5; DB 21; Length 371;
 Best Local Similarity 29.7%; Pred. No. 2.6e-28;
 Matches 83; Conservative 60; Mismatches 111; Indels 25; Gaps 7;
 QY 34 TERKEALLTVTKANLRNTM---RTLHYIIGHGIPLYRFSSSIVPLATHPDVMDVFTPF- 89
 Db 48 tigrdglesvkqlgtqnvldliklvevnhnfhgmrvssdlfpfash--akygytlefa 105
 QY 90 QKEFREIGELVKTHTLTSFHPNOFTLTSPKESVTKNAVTDMAHYHMLEAMGIADR-- 147
 Db 106 qshleevgklankynhrhltmhpggytqiasprevvvdsairldayhdeilsrmkneqln 165
 QY 148 --SVINIHIGGAYGNKDTATAQFHQNIKQLPQEIKERMTLENDDKTYTTEPLQVCEQED 205
 Db 166 kdavliihlggtfegkktldfrknyqrlsdsvkarivlendddvswsqdlplcqln 225
 QY 206 VPEVDFDHFHYANP----DOHADLNVALPRMIKTWERIGLQPKVHLSPPKSEQAI----- 256
 Db 226 iplvldwhhhniavpgtlregslldmpliretctwrgitkgkqghysesadptaisgmkr 285
 QY 257 RSHADYVDANFLERFROWGTNIDFMIEAKOKKALLRL 295
 Db 286 rahsdrv-----fdfpccptmdlmieakeqavfel 318
 RESULT 2
 AAY44500
 ID AAY44500 standard; Protein: 600 AA.
 XX
 AC AAY44500;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE GST signal peptide and delta228 S. pombe UVDE fusion protein.
 KW GST signal peptide; glutathione-S-transferase signal peptide;
 KW ultraviolet damage endonuclease; UVDE; delta228-UVDE;
 KW repair-deficient E. coli strain; UV irradiation; DNA damage;
 KW UV radiation damage; photoproduct; abasic site; aplatium diaduct;
 KW mismatched nucleotide pairing; nucleotide alkylation;
 KW fusion protein; skin cancer.
 OS Schizosaccharomyces pombe.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 11 /note= "Encoded by aaaa"
 FT Misc-difference 85 /note= "Encoded by gtt"
 FT Misc-difference 147 /note= "Encoded by gtt"
 XX
 PN W09963828-A1.

XX
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX
 PA (UYEM-) UNIV EMORY.
 XX
 PI Doetsch PW, Kaur B, Avery AM;
 XX WPI; 2000-116417/10.
 DR N-PSDB; AAZ29860.
 XX
 PT A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers -
 XX
 PS Disclosure; Page 56; 133pp; English.
 XX
 CC The present sequence is a fusion protein comprising the GST signal
 CC peptide and delta228 truncated ultraviolet damage endonuclease (UVDE)
 CC from S. pombe. Delta228-UVDE is a 288 amino acid deletion of the
 CC N-terminal of the S. pombe uvel+ gene product. This is expressed in frame
 CC with a GST leader sequence. Stable endonuclease fragments can be produced
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, aplatium diaduct, an intercalated
 CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 XX Sequence 600 AA;
 SQ
 Query Match 21.6%; Score 359.5; DB 21; Length 600;
 Best Local Similarity 29.7%; Pred. No. 5.5e-28;
 Matches 83; Conservative 60; Mismatches 111; Indels 25; Gaps 7;
 QY 34 TERKEALLTVTKANLRNTM---RTLHYIIGHGIPLYRFSSSIVPLATHPDVMDVFTPF- 89
 Db 277 tigrdglesvkqlgtqnvldliklvevnhnfhgmrvssdlfpfash--akygytlefa 334
 QY 90 QKEFREIGELVKTHTLTSFHPNOFTLTSPKESVTKNAVTDMAHYHMLEAMGIADR-- 147
 Db 335 qshleevgklankynhrhltmhpggytqiasprevvvdsairldayhdeilsrmkneqln 394
 QY 148 --SVINIHIGGAYGNKDTATAQFHQNIKQLPQEIKERMTLENDDKTYTTEPLQVCEQED 205
 Db 395 kdavliihlggtfegkktldfrknyqrlsdsvkarivlendddvswsqdlplcqln 454
 QY 206 VPEVDFDHFHYANP----DOHADLNVALPRMIKTWERIGLQPKVHLSPPKSEQAI----- 256
 Db 455 iplvldwhhhniavpgtlregslldmpliretctwrgitkgkqghysesadptaisgmkr 514
 QY 257 RSHADYVDANFLERFROWGTNIDFMIEAKOKKALLRL 295
 Db 515 rahsdrv-----fdfpccptmdlmieakeqavfel 547
 RESULT 3
 AAY44498
 ID AAY44498 standard; Protein: 828 AA.
 XX
 AC AAY44498;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE GST signal peptide and S. pombe UVDE fusion protein.
 KW GST signal peptide; ultraviolet damage endonuclease; UVDE; Uvelp;
 KW UV irradiation; DNA damage; UV radiation damage; fusion protein;
 KW skin cancer; glutathione-S-transferase.

```

XX OS Schizosaccharomyces pombe.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT Peptide 1..229
XX FT /label= GST_signal_peptide
XX FT Protein 230..828
XX FT /note= "S. pombe UVDE"
XX FT Misc-difference 11
XX FT /note= "Encoded by aaaa"
XX FT Misc-difference 85
XX FT /note= "Encoded by gtt"
XX FT Misc-difference 147
XX FT /note= "Encoded by gtt"
XX PN W09963828-A1.
XX PD 16-DEC-1999.
XX PD 08-JUN-1999; 99WO-US12910.
XX PF 08-JUN-1998; 98US-0088521.
XX PR 18-MAY-1999; 99US-0134752.
XX PA (UYEM-) UNIV EMORY.
XX PI Doetsch PW, Kaur B, Avery AM,
XX PX WPI; 2000-116417/10.
XX DR N-PSDB; AAZ29858.
XX A new truncated ultraviolet damage endonuclease for treatment of skin
XX PT cancers.
XX PS Claim 7; Page 51; 133pp; English.
XX CC The present sequence is a fusion protein comprising the GST signal
XX CC peptide and full length ultraviolet damage endonuclease (UVDE) from
XX CC Schizosaccharomyces pombe. UVDE genes are isolated by introducing a
XX CC foreign cDNA library into a repair-deficient E. coli strain and
XX CC selecting for complemented cells by UV irradiation of the transformants.
XX CC This provides stable endonuclease fragments for cleaving a double-
XX CC stranded DNA molecule that has a distorted structure resulting from UV
XX CC radiation damage, a photoproduct, an abasic site, mismatched nucleotide
XX CC pairing, apurination diaduct, an intercalated molecule or alkylation of a
XX CC nucleotide. Uvdp can be used in compositions for internal or topical
XX CC application and as a therapeutic agent for skin cancers.
XX SQ Sequence 828 AA;

Query Match 21.6%; Score 359.5; DB 21; Length 828;
Best Local Similarity 29.7%; Pred. No. 8.9e-28;
Matches 83; Conservative 60; Mismatches 111; Indels 25; Gaps 7;

QY 34 TERKALLTVTKANLRNTM---RTLHYIIGHGIPLYRFSSTIVPLATHPDVWMDVFTVPF- 89
DB 505 ttiqrglesvkgltqgnvldikivewnhnfgihfmrvsddlfafsh--akygtylafa 562
QY 90 QKEFREIGELVKTHOLRFSHPNQFTLTSPKESVTKNAVTDMAHYRMLEAMGIADR-- 147
DB 563 qshleevgklankynhrithmpggytqiassprevvdsairdlayhdeilsrmlneqln 622
QY 148 --SVNIIHIGAYGNKDPATQAFQHNKQLPQETKERMTLENDOKTYYTTEETLOVCEQED 205
DB 623 kdavililigtfeqkkteldrfrknygridsvkarlivlenddsvsqvdlplcqln 682
QY 206 VPFVDFHHFYANP-----DDHADLNVALPRMIKTWERTIGLPKVLHSSPKSEQAI----- 256
DB 683 iplvldwhhniavgtrregaldmptiretwtkrktgkqkhyesadptalsgmkr 742
QY 257 RSHADYVDANFLLERFRQWGTNIDFMIEAKOKKALLRL 295

Query Match 20.8%; Score 345; DB 17; Length 656;
Best Local Similarity 32.3%; Pred. No. 1.9e-26;
Matches 87; Conservative 51; Mismatches 105; Indels 26; Gaps 8;

QY 46 ANLRNTMRTLHYIIGHGIPLYRFSSTIVPLATHPDVWMDVFTVPFQKE-FREIGELVKTHQ 104
DB 277 anardivkmclwckeygtrfirlsemfpfashp-vhgyklapfasevlaeagrvaalg 335
QY 105 LRTSFHPNQFTLTSPKESVTKNAVTDMAHYRMLEAMGIADR---SVNIIHIGAYGN 160
DB 336 hrltthpgqftqlgsprkevesairdleyhdellsllklpedqndrdavmllmngqfgd 395
QY 161 KQATQAFQHNKQLPQETKERMTLENDOKTYYTTEETLOVCEQEDVPFVDFHHFYANPD 220
DB 396 kaatlerfrknyarsqsknrilvleddvgwtvhdllpvcceelnipmvlvdyhhnicfd 455
QY 221 D-----HADLNVALP-----RMIKTWERTIGLPKVLHSSPKSEQAI-----RSHADYVDANF 267
DB 456 pahlregtdlisdgpkqlqeriantwtkrktgkqkhysep-cdgavtprhrkrhrprv---- 510
QY 268 LLERFRQWGTNIDFMIEAKOKKALLRLM 296

```

Db 511 --mtlppcpdmdlmieakdkegavfelm 537

RESULT 5
ID AAY44502 standard; Protein; 626 AA.
XX AC AAY44502;
XX DT 27-MAR-2000 (first entry)
XX DE N. crassa delta228-UV damage endonuclease.
XX DE Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
KW glutathione-S-transferase signal peptide; uvel+ gene product;
KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
KW abasic site; platinum diaduct; mismatched nucleotide pairing;
KW nucleotide alkylation; skin cancer.
XX OS Neurospora crassa.
XX PN WO963828-AL.
XX PD 16-DEC-1999.
XX PF 08-JUN-1999; 99WO-US12910.
XX PR 08-JUN-1998; 98US-0088521.
XX PR 18-MAY-1999; 99US-0134752.
XX PA (UYEM-) UNIV EMORY.
XX PI Doetsch PW, Kaur B, Avery AM;
XX WPI; 2000-116417/10.
XX DR A new truncated ultraviolet damage endonuclease for treatment of skin
XX PT cancers -
XX PS Claim 16; Page 59; 133pp; English.
XX CC The present sequence is delta228-UV damage endonuclease from N. crassa.
CC Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
CC uvel+ gene product. This is expressed in frame with a GST leader sequence
CC to generate a fusion protein. This provides stable endonuclease fragments
CC for cleaving a double-stranded DNA molecule that has a distorted
CC structure resulting from UV radiation damage, a photoproduct, an abasic
CC site, mismatched nucleotide pairing, platinum diaduct, an intercalated
CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
CC for internal or topical application and as a therapeutic agent for skin
XX CC cancers.
XX SQ Sequence 626 AA;
Query Match 18.9%; Score 313.5; DB 21; Length 626;
Best Local Similarity 30.5%; Pred. No. 3.le-23;
Matches 82; Conservative 49; Mismatches 97; Indels 41; Gaps 9;
Qy 46 ANLRNTRTLHYIIGHGTPLYRFSSSIVPLATHPDVMDVFTPFQKE-FREIGELVKTHQ 104
Db 277 anardivkmclwf-----pfashp-vhgyklapfasevlaeagrvaalg 320
Qy 105 LRTSEHPNQFTLFTSPKESVTKNAVDMAYHYRMLAMGIADR----SVINIHIGGAYGN 160
Db 321 hrlthpqgftqlgsprkvesairldleyhdllslkipeqgnrdavmlhmgsqfgd 380
Qy 161 KDTATAQPHQNIKQLPOEIKERTMTLENDKTYTTEFTLQVCEQEDVPFDFHHFYANPD 220
Db 381 kaatlerfknyarlsqscnrlvlenddvgtvhdilpvcceinlpmvldyhnhncifd 440
Qy 221 D-----HADNLNALP----RMKLTWERGLCPKPHLSLSPKSEQAI-----RSHADYVDANF 267


```

Db      561 elge--dsqmlaheyypr-----skmsrslksamkd---hynfitayenllfdadvidn 611
Qy      152 -----THIGGAYGNKDTATQAQFQNIKOLPO-EIKERMTLENDKTY-----TTEET- 197
Db      612 dagkqfinlagvntspdgaaantvwmkskrtpetpynllhlinlvndqnrnsngnqtaqtn 671
Qy      198 ----LOVCEQEDVPFVDFHHFYANPDHDLNVALP-----RMIKTWE 237
Db      672 latkvyigaetigv-----yaasphngatqslpftgttdssgsyisftvpsleyws 726
Qy      238 RIGLOPKVHLSSP-----KSEQATRS-----HADVVDANFLLEFRONGTNIIDFMIEA 285
Db      727 mlym--krstaapvdmnyeataiksnsvntnhagytgsqf-vdgfatvndgvsfiiva 783
Qy      286 KQDKKALLR 294
Db      784 skddyvlr 792

RESULT 9
AAW97216
ID AAW97216 standard; Protein; 1040 AA.
XX
AC AAW97216;
XX
XX 06-MAY-1999 (first entry)
XX
DE An isoleucyl tRNA (ileS) synthetase.
XX
XX Isoleucyl tRNA synthetase; ileS; vaccine; ocular; genital trachoma;
KW lymphogranuloma venereum; keratitis; corneal infiltration; prostaticitis;
KW bacterial adhesion; in-dwelling device; cell invasion.
XX
OS Chlamydia trachomatis.
XX
XX Key Location/Qualifiers
XX
XX Misc-difference 1 /note= "encoded by GTG"
XX
XX EP893495-A2.
XX
XX 27-JAN-1999.
XX
XX 10-JUL-1998; 98EP-0305502.
XX
XX 23-JUL-1997; 97US-0898798.
XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX (SMIK ) SMITHKLINE BEECHAM PLC.
XX
XX Brown JR, Lawlor EJ, Reichard R;
XX
XX WPI; 1999-097780/09.
XX N-PSDB; AAX15503.
XX
XX New Chlamydia trachomatis isoleucyl tRNA synthetase polypeptide and
PT polynucleotide - useful as diagnostic reagents and for prevention
PT and treatment of Chlamydia infections which cause ocular trachoma
PT and infertility
XX
XX Claim 6; Page 9-10; 34pp; English.
XX
XX The present sequence represents an isoleucyl tRNA (ileS) synthetase
CC polypeptide. Iles polypeptides are useful for screening for specific
CC inhibitors or activators, in vaccines to induce an immunological
CC response, to raise antibodies and for treating conditions associated
CC with a lack of iles protein. Iles antagonists are used to treat
CC conditions requiring inhibition of iles, specifically infection
CC by C. trachomatis, including ocular or genital trachoma, lymphogranuloma
CC venereum, keratitis, corneal infiltration and prostaticitis etc. Iles
CC polypeptides, polynucleotides and their antagonists can also be used to
CC prevent adhesion of bacteria, especially Gram-positive, to extracellular

```

```

CC matrix proteins in wounds or on in-dwelling devices, and to block iles
CC polypeptide-mediated cell invasion or normal progression of infection.
XX
SQ Sequence 1040 AA;

```

Query Match 5.6%; Score 93.5; DB 20; Length 1040;

Best Local Similarity 19.7%; Pred.No.2.6; 104; Indels 87; Gaps 13;
Matches 57; Conservative 42; Mismatches

```

Qy 71 SIIVPLATHPDVMM-----DFVT-----PFQKFEIGEIVKTHQLRSTFHP 111
   | : | | | | : : : | : | : | : | : | : | : | : | : | : | : |
Db 234 smmalavhellyvrikdkesdevylqeslprwfpdresyewigqlsgksalvqsyep 293
   | : | | | | : : : | : | : | : | : | : | : | : | : | : | : |
Qy 112 NQFTLTSPKESVTNNAVTDMAHYHRYMLEAMGIADRSVIN-----THIGGAYGNKDTA 164
   | : | | | | : : : | : | : | : | : | : | : | : | : | : | : |
Db 294 -lfpyfqdkke-----leafrilpadfiesegtgivhmapafgeadff 336
   | : | | | | : : : | : | : | : | : | : | : | : | : | : | : |
Qy 165 TAQFH-----QNIKOLPQEIKERMTLENDKTY-----TTEETL 198
   | : | | | | : : : | : | : | : | : | : | : | : | : | : | : |
Db 337 acqehnvplvcvndggcytaevkdfvgeyiksaukgiarr--lknenklfyqgtvrhry 394
   | : | | | | : : : | : | : | : | : | : | : | : | : | : | : |
Qy 199 QVCEQEDVPFVDFHHFYANPDHDLNVALPRMIKTWERIGLOPKVHLSSPKSQAIRS 258
   | : | | | | : : : | : | : | : | : | : | : | : | : | : | : |
Db 395 pfcwrtdspilykavns-----fvavekkskmlkanesihwtpe-hikgrgfrgkwieg 448
   | : | | | | : : : | : | : | : | : | : | : | : | : | : | : |
Qy 259 HADYVDANFLLEFRONGTNIIDFMIEAKOKKALLRL--MDELSSIRGVK 306
   | : | | | | : : : | : | : | : | : | : | : | : | : | : | : |
Db 449 ardwa-----isrnywgtpipiw-----rsddqellivigsiglealsgqk 490
   | : | | | | : : : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 10

AAW56782
ID AAW56782 standard; Protein; 1041 AA.

XX AAW56782;

XX 27-MAR-2000 (first entry)

XX C. trachomatis iles polypeptide.

XX Isoleucyl tRNA synthetase; ileS; antibiotic; antibacterial; vaccine;
KW ocular trachoma; lymphogranuloma venereum; inclusion conjunctivitis;
KW mucopurulent rhinitis; urethritis; tropical bubo; cervical follicle;
KW prostaticitis; proctitis; infant pneumonitis.

XX Chlamydia trachomatis.

XX US6001602-A.

XX 14-DEC-1999.

XX 23-JUL-1997; 97US-0898978.

XX 23-JUL-1997; 97US-0898978.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Brown JR, Lawlor EJ, Reichard RW;

XX WPI; 2000-096135/08.

XX N-PSDB; AAZ46652.

XX Nucleic acids encoding Chlamydia trachomatis isoleucyl tRNA synthetase
PT polypeptides useful for treating, for example, urethritis, proctitis
PT and prostaticitis -

XX Claim 1; Columns 9-11; 21pp; English.

XX This represents a isoleucyl tRNA synthetase (ileS) polypeptide derived
CC from Chlamydia trachomatis. The iles polypeptide can be expressed by
CC standard recombinant methodology. The protein produced may then be used
CC to screen for antibiotics active against C. trachomatis and to study the

CC role of the iles protein in the pathogenesis of disease. They are
 CC particularly useful as antigens in the production of antibodies specific
 CC for C. trachomatis. The antibiotics and antibodies obtained are useful
 CC for treating a wide range of diseases including classic ocular trachoma,
 CC lymphogranuloma venereum, inclusion conjunctivitis, mucopurulent
 CC rhinitis, urethritis, tropical bubo, cervical follicles, prostatitis,
 CC proctitis and infant pneumonitis. The antibodies may also be used to
 CC detect the presence of C. trachomatis and related microbes according to
 CC standard assay protocols.

XX Sequence 1041 AA;

Query Match 5.6%; Score 93.5; DB 21; Length 1041;

Best Local Similarity 19.7%; Pred. No. 2.6; Mismatches 42; Indels 87; Gaps 13;
 Matches 57; Conservative 42;

Qy 71 SIVPLAPHDPMV-----DFVT-----PFOKEFREIGELVKTHOLRTSFHP 111
 | : | | :
 Db 234 snmalavhpeltyvrikkdesgdeylilgqeslprwfpdresyewiglsksgivgsgsyep 293
 Qy 112 NQFTLFTSPKESYTKNAVDMAYHRYMLEAMGADRSVIN-----IHGGAYGNKDTA 164
 | : | | :
 Db 294 -lfpyfqdkke-----leafriipadfielsegtgihmapafgeadff 336
 Qy 165 TAQPH-----ONIKQLPOEIKERMTELENDKTY-----TTEETL 198
 | : | | :
 Db 337 acqehnvplvcvdpngqcytaevkdfvgeyiksadgiarr--lknenklfyggtvrhry 394
 Qy 199 QVCEQEDVPVFDFHFFYANPDHADLNVALPRMIKTWIRIGLPKVLHSSPKSEQAIRS 258
 | : | | :
 Db 395 pfcwrttdsplykavns-----fvavekvksmlkanesihwtpe-hikqgrfgkwleg 448
 Qy 259 HADYVDANELLERPRQGTNIDFIEAKQKDKALLRL--MDELSIRGVK 306
 | : | : | | :
 Db 449 ardwa-----lsrnywgtptipw---rsddgellvigsqealealsgqk 490

RESULT 11

AAB03759
 ID AAB03759 standard; Protein; 234 AA.

AC AAB03759;

DT 04-OCT-2000 (first entry)

DE Human H37 amino acid sequence #2.

XX H37; human; Cdc7 regulatory subunit; cytostatic; proliferative; cancer;
 KW anti-proliferative; replication regulator; stem cell.

OS Homo sapiens.

XX WO200026250-A1.

XX 11-MAY-2000.

XX 01-NOV-1999; 99WO-JP06076.

XX 30-OCT-1998; 98JP-0311408.

XX (NTSC-) JAPAN SCI & TECHNOLOGY CORP.

PA (ARAI/) ARAI K.

PA (MASAI/) MASAI H.

PI Arai K, Masai H;

XX WPI; 2000-365580/31.

XX N-PSDB; AAA53484.

Human H37 proteins with a Cdc7 activity regulatory subunit, for
 controlling cell replication and cell proliferation, useful in treating
 cancers and diseases due to abnormal production of stem cells

XX Claim 2; Page 46-47; 55pp; Japanese.

XX The present sequence represents a human H37 protein sequence. H37 is a
 CC protein with a Cdc7 activity regulatory subunit. The invention relates to
 CC two H37 protein and nucleotide sequences. H37 proteins exhibit
 CC cytostatic, proliferative, anti-proliferative, and cell replication
 CC regulatory activities. The proteins, encoded genes and DNA fragments are
 CC useful in treating cancers and other diseases resulting from abnormal
 CC production of stem cells. Antibodies directed against one of the H37
 CC proteins can be used to inhibit cell proliferation.

XX Sequence 234 AA;

Query Match 5.6%; Score 92.5; DB 21; Length 234;

Best Local Similarity 31.7%; Pred. No. 0.33;
 Matches 33; Conservative 10; Mismatches 42; Indels 19; Gaps 4;

Qy 227 VALPRMIKTWIRIGLPKVLHSSPKS-----EQAIRSHADYVDANFLLEFR 273
 | : | | : : : : : : | | | | : : : : : : : : : : : : : : : :
 Db 105 vspesaytaetshpshdsgsfkspdtvelsrqkllvekaikdh-dfipsnlsnal 163

Qy 274 QMGNTNI-----DFMIEAKQKDKALLRLMDELSSIR-GVXRIGGGA 312
 | : | | : : : : : : | | | | : : | | : : | | : : | | : : | |

Db 164 swgvkilhiddirryieqkkkelyllkksstsvrdggkrvgsga 207

RESULT 12

AAB42482

ID AAB42482 standard; Protein; 674 AA.

AC AAB42482;

XX 08-FEB-2001 (first entry)

DE Human ORFX ORF2246 polypeptide sequence SEQ ID NO:4492.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnerary; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.

OS Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI; 2000-602362/57.

XX N-PSDB; AAC76691.

PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX Claim 11; Page 3676-3677; 5507pp; English.
 PS AACT74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
 CC antiproliferative; antiparkinsonian; nootropic; immunosuppressive;
 CC osteoprotective; anticoagulant; antithrombotic; neuroprotective;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX Sequence 674 AA;

Query Match 5.6%; Score 92.5; DB 21; Length 674;
 Best Local Similarity 31.7%; Pred. No. 1.7;
 Matches 33; Conservative 10; Mismatches 42; Indels 19; Gaps 4;
 QY 227 VALPRMIKTWIRIGLOPKVHLSPKS-----EQAIRSHADYVDANFLRLER 273
 Db 105 vspesaytaettspshdsgsfkspdtvclsrqkllvekaikdh-dfipsnlsnal 163
 QY 274 QWGTNI-----DFMTEAKOKKALLRLMDELSSIR-GVKRIGGGA 312
 Db 164 swgkvlhiddirryieqkkkellyllkksstsvrdggrvgsa 207

RESULT 13
 AAB03758
 ID AAB03758 standard; Protein; 674 AA.
 XX AAB03758;
 XX 04-OCT-2000 (first entry)
 XX Human H37 amino acid sequence #1.
 XX H37; human; Cdc7 regulatory subunit; cytostatic; proliferative; cancer;
 XX anti-proliferative; replication regulator; stem cell.
 XX Homo sapiens.
 XX WO200026250-A1.
 XX 11-MAY-2000.
 XX 01-NOV-1999; 99WO-JP06076.
 XX 30-OCT-1998; 98JP-0311408.
 XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
 XX (ARAI/) ARAI K.
 XX (MASA/) MASAI H.
 XX Arai K, Masai H;
 XX

DR WPL; 2000-365580/31.
 DR N-PSDB; AAB53483.
 XX Human H37 proteins with a Cdc7 activity regulatory subunit, for
 PT controlling cell replication and cell proliferation, useful in treating
 PT cancers and diseases due to abnormal production of stem cells -
 PS Claim 1; Fig 5; 55pp; Japanese.
 XX The present sequence represents a human H37 protein sequence. H37 is a
 CC protein with a Cdc7 activity regulatory subunit. The invention relates to
 CC two H37 protein and nucleotide sequences. H37 proteins exhibit
 CC cytostatic, proliferative, anti-proliferative, and cell replication
 CC regulatory activities. The proteins, encoded genes and DNA fragments are
 CC useful in treating cancers and other diseases resulting from abnormal
 CC production of stem cells. Antibodies directed against one of the H37
 CC proteins can be used to inhibit cell proliferation.
 XX Sequence 674 AA;

Query Match 5.6%; Score 92.5; DB 21; Length 674;
 Best Local Similarity 31.7%; Pred. No. 1.7;
 Matches 33; Conservative 10; Mismatches 42; Indels 19; Gaps 4;
 QY 227 VALPRMIKTWIRIGLOPKVHLSPKS-----EQAIRSHADYVDANFLRLER 273
 Db 105 vspesaytaettspshdsgsfkspdtvclsrqkllvekaikdh-dfipsnlsnal 163
 QY 274 QWGTNI-----DFMTEAKOKKALLRLMDELSSIR-GVKRIGGGA 312
 Db 164 swgkvlhiddirryieqkkkellyllkksstsvrdggrvgsa 207

RESULT 14
 AAW77094
 ID AAW77094 standard; Protein; 522 AA.
 XX AAW77094;
 XX 09-NOV-1998 (first entry)
 XX Sorting nexin 1.
 XX Human; sorting nexin; intracellular domain; cell surface receptor;
 XX translocation; lysosome; down-regulation; cancer.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FT Modified-site 115
 FT /note= "Xaa not specified, encoded by TTT"
 XX US5804412-A.
 XX 08-SEP-1998.
 XX 01-APR-1996; 96US-0625322.
 XX 01-APR-1996; 96US-0625322.
 XX (REGC) UNIV CALIFORNIA.
 XX Cadena DL, Gill GN, Kurten RC;
 XX WPI; 1998-505652/43.
 XX N-PSDB; AAV48248.
 XX Nucleic acids encoding sorting nexin proteins - used for recombinant
 XX production and regulating expression of the protein, especially in
 XX treating unregulated cell growth e.g. cancer
 PS Claim 3; Column 23-28; 25pp; English.

XX Sorting nexins (SNX) have the ability to bind an intracellular domain of
 CC an activated cell surface receptor expressed on a cell and direct
 CC translocation of the receptor to a lysosome in the cell. The nucleic
 CC acids encoding SNX are useful for the down-regulation of cell surface
 CC receptors for degradation when they are no longer required to carry out a
 CC signal. The nucleic acids can be used to express the protein in a cell
 CC where expression is desired e.g. where cancer is caused by up-regulation
 CC of receptors. Antisense constructs can be used to inhibit native
 CC expression of the protein.
 XX Sequence 522 AA;

Query Match 5.5%; Score 91.5; DB 19; Length 522;
 Best Local Similarity 19.3%; Pred. No. 1.4;
 Matches 64; Conservative 54; Mismatches 106; Indels 107; Gaps 16;

QY 28 YSKLSKTER-----KEALLTVTKANL-RNTWRTLHYIIGHGIPLYRFSSSIVPLA 76
 | ||| :||: || :
 Db 194 yeklsekhsqngfiivpppeksligmtkvkgkedsssaeflekrraalerylqriv--- 250
 QY 77 THDDVMDFVTPQKFEIGELVKTHQLTSFHPNQFLTSPKESVTKNAV-----TDM 132
 | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
 Db 251 nhptmlqg---pdvreflekeelpavgvtqlsgagllmfknkatdavsakmtikmnesdi 307
 QY 133 AYHYRMLEANGIADR-----SVINIHIGGAYGNKDTA--TAOPHONIKOLPOEIKERMTL 185
 : : : | :
 Db 308 wfeeklqeveceeqrlrkhlhavvetlrvn---hrkelalntaqfaksiamlg----- 355
 QY 186 ENDDKVTTEETLQVCEQEDVPFVDFHFHYANPD-----DH-----ADLNVALPRM 232
 : : | : | : : : | : | : | : | : | : | : | : | : | : | : | :
 Db 356 sseentalsralsqlaeveek---ieqlhqeanndfllaellsdyirllaivraafqdr 413
 QY 233 IKTWERTGLQPKVHLSSPKSEQAIRSHADYVDANFILLERPRQ-----NCTNIDFMIEAK 286
 : ||| :
 Db 414 mktwqr-----wqdaqatlgkreaearllwankpdklqqak 450
 QY 287 QKOKALLRLMDELSSIRGVKRIGGGALOWKS 317
 ||| :
 Db 451 -----del-----lews 458

RESULT 15
 AAY77979
 ID AAY77979 standard; protein; 558 AA.
 AC AAY77979;
 XX
 XX 14-JUN-2000 (first entry)
 XX
 DE A. thaliana environmental stress tolerance related protein.
 XX
 XX Environmental stress; plant; transgenic plant; anaerobic; flooding; cold;
 KW dehydration; drought; heat stress; salinity; osmotolerance.
 XX Arabidopsis thaliana.
 OS
 XX W0200008187-A2.
 XX
 XX 17-FEB-2000.
 XX
 XX 04-AUG-1999; 99WO-EP05652.
 PF
 XX 04-AUG-1998; 98EP-0202634.
 PR
 XX (VLAAs-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
 PA
 XX Lee JH, Verbruggen N;
 PI
 XX WPI; 2000-205726/18.
 DR
 XX N-PSDB; AAZ98359.
 XX

PT Isolation of polynucleic acids useful for producing transgenic plant by
 PT isolating genes involved in tolerance to environmental stress -
 XX Claim 12; Page 278-282; 312pp; English.
 PS
 XX The invention relates to isolation of coding sequences and/or genes
 CC involved in tolerance to environmental stress in plants. The sequences
 CC (AAZ98305-298365) are useful for producing a transgenic plant having
 CC enhanced tolerance or resistance to environmental stress conditions such
 CC as anaerobic, flooding, cold, dehydration, drought, heat stress or
 CC salinity. This is useful for producing improved yield, growth,
 CC development and productivity under environmental stress conditions, and
 CC also provides growth of crops in areas where they cannot grow without
 CC the induced osmotolerance. Sequences AAY77925-984 represent polypeptide
 CC sequences from A. thaliana that are encoded by the genes involved in
 CC environmental stress tolerance.
 XX Sequence 558 AA;

Query Match 5.5%; Score 91; DB 21; Length 558;
 Best Local Similarity 21.8%; Pred. No. 1.8;
 Matches 53; Conservative 45; Mismatches 91; Indels 54; Gaps 11;

QY 69 SSSIPLATHPDV-----MMDFVT--PFOKEFREIGELVKTHQLTSFHPNQFLTFS 119
 | | : : | : | : | : | : | : | : | : | : | : | : | : | :
 Db 114 srsrvssksnptvdafqgkewekltadpgtrvyleqddfvktmk-eiqqrpnlnlymk 172
 QY 120 PRESVTKNAVTDMAHYHYRMLEAMGIADRSVINIHIGGAYGNKDTATATQFHQNIKOLPOEI 179
 | : : | : : | : : | : : | : : | : : | : : | : : | : : | : :
 Db 173 dk-----rvmkalgv-----lrvkfvgssg-edtemkadeerkepepeme 212
 QY 180 KERMTLENDKTTYTTEETLQVCEQEDVPF-VPDF-----HHFYANPDHDLNVALPRMI 233
 : : | : | : | : | : | : | : | : | : | : | : | : | : | : :
 Db 213 pmeleeeerqkkrkekalkkegnvaykkdkdgravehytkamelddedisylnr-- 270
 QY 234 KTWERTGLQPKVHLSSPKSEQAIRSHADYVDANFILLERFROMGTNIDFMIEA-KOKDKAL 292
 : : | : | : | : | : | : | : | : | : | : | : | : | : | : :
 Db 271 -----aavylemgkyecie-----dcdkavergreirsdfkmiarlttrkgsal 315
 QY 293 LRL 295
 : : :
 Db 316 vkm 318

Search completed: January 15, 2002, 13:50:09
 Job time: 298 sec


```

FT REPEAT 1366 1477 4.
FT REPEAT 1487 1598 5.
FT REPEAT 1608 1720 6.
FT REPEAT 1730 1842 7.
FT REPEAT 1851 1964 8.
FT REPEAT 1974 2086 9.
FT REPEAT 2096 2204 10.
FT REPEAT 2214 2326 11.
FT REPEAT 2335 2447 12.
FT REPEAT 2457 2569 13.
FT REPEAT 2579 2689 14.
FT REPEAT 2699 2808 15.
FT REPEAT 2818 2928 16.
FT NP_BIND 3034 3041 ATP (POTENTIAL).
FT VARSPLIC 136 495 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 3256 AA; 358741 MW; 578F8C51BED42517 CRC64;

Query Match 3.98; Score 137; DB 1; Length 3256;
Best Local Similarity 19.98; Pred. No. 2.4;
Matches 146; Conservative 92; Mismatches 296; Indels 200; Gaps 30;

Qy 5 KSKAAALD-----TP-QSESTFSSLDSSAPSA-----RNLRRSGRNLIQPS 48
Db 2428 KEKAEALDVGFKELQTPCHTEESMTDDKITEVSKSPQSPSEFKTSKQRLKIPLV 2487
Qy 49 EKDRDHEKRSGEELAGRMGMKDANGCHLREG-----KEQEGVKMAIEGLARME-RLQR 102
Db 2488 KVMKEEPLAYSKLIT-RTSGETTQHTPTGTDGSKSIKAFKESPKQIILDPAAASVTGSRRL 2546
Qy 103 ATKROKKOLEDDGIPVPSVSKRFTAPVHHKSTNAEREAKEPV----- 146
Db 2547 RTRKEKALELDVDFKELFS-----AGHTEESMTIDKNTKIPCKSPPELTDTATSKR 2602
Qy 147 -----LKTHSKDVERAEI-GVDDVVKMEPA-----TNII-----EPEDAQAAE 186
Db 2603 CPKTRPRKEVEELSASVRLTQSTHTHKEPASGDEGIKVLKQRAKKKPNVEEPS 2662
Qy 187 RGAARPPAVNSYLLPLWPKGRGLVACLYLNKAKPTIFFSRTCRMASIVDHRPIQFED 246
Db 2663 RRRPRAPKEAQ-----PLEDLAGEFTEUSETSGHTQESLTAGKATK-----IPCESPLEVVD 2715
Qy 247 ---EPEHLKPKDKSKPEQDGLGHKFQVELGLANARDIVKMLCWNKYGIRFLRLSSE 303
Db 2716 TTASTKRHLRTVQKVQKEEPSAVKFTQSG--ETTDADREPAGEDK-GIKALKESAKQ 2772
Qy 304 FFPASHPVHGYKLAPFASEVLAE-----AGRVAAELGH-----RLTHTP-----QQFT 346
Db 2773 TPAPAAVSGTSSRRRPRAPRESAQAIEDLAGFKDPAAGHTTEESMTDDKTKIPCKSSPELE 2832
Qy 347 QLGSPRKEVESAIIRDLEVHDELLSLKL-----PQQN 380
Db 2833 DTATSKRRPRTRAQKVEKVELLAVGLKLTQSTGETTHDKPEVGEKGTKAFKQAPAKRN 2892
Qy 381 RDVAMIIHMGGO---FGDKAATLERFKRNYARLSQSKCNRLVLENDVGVTVHDLPLVCE 437
Db 2893 VDAEDVIGSRQRPRAPKEKAQPLEDL-ASFQELSOT----- 2927
Qy 438 ELNIPWLDYHHNICFOPAHLE-----GTLDISDP-KLOERIANTWKRK 482
Db 2928 -----PGHTEELANGAASFTSAPKQTPDQSGPLKISRVLRAPK-- 2967
Qy 483 GIKQKMHYSEPCDGAVTPRDRKRHRPRVMT-LPPCPPDMDLMIEAKDKEQAVFELMRTEK 541
Db 2968 -----VEPVGDVVSTRDPVKSKSNTSLPPLP-----FKRGGGKGDSVTGTGKRLRC 3014
Qy 542 LCGFEKINDWYDRDDENRPPAPKAKKKKGGKRRKTTDEEAPEPEVDTAADDVKDA 601
Db 3015 MPAPAEIIEELPASKQKVAPRARGKSPEPVVIMKRSRLTSAKRIEPAE-ELNSNDMKTN 3073
Qy 602 PEGPK---EVPPEE 612
Db 3074 KEHKLODSVPENK 3087

```

```

RESULT 5
NFM_MOUSE
ID NFM_MOUSE STANDARD: PRT: 848 AA.
AC P08553; Q61961;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUROFILAMENT TRIPLET M PROTEIN (160 KDA NEUROFILAMENT PROTEIN)
DE (NEUROFILAMENT MEDIUM POLYPEPTIDE) (NF-M).
GN NEF3 OR NEFM OR NFM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87246694; PubMed=3036526;
RA Levy E., Liem R.K.H., D'Eustachio P., Cowan N.J.;
RT "Structure and evolutionary origin of the gene encoding mouse NF-M,
RT the middle-molecular-mass neurofilament protein.";
RL Eur. J. Biochem. 166:71-77(1987).
RN [2]
RP SEQUENCE OF 322-540 FROM N.A.
RX MEDLINE=87158637; PubMed=3103856;
RA Julien J.-P., Meyer D., Flavell D., Hurst J., Grosveld F.;
RT "Cloning and developmental expression of the murine neurofilament
RT gene family.";
RL Brain Res. 387:243-250(1986).
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NEM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NEF RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05640; AAA29127.1; -.
DR EMBL; M20481; AAA39815.1; -.
DR PIR; S00030; S00030.
DR MGD; MGI:97314; Nfm.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; Ifilament; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW Glycoprotein.
FT INIT_MET 0
FT DOMAIN 1 102 HEAD.
FT DOMAIN 103 410 ROD.
FT DOMAIN 411 848 TAIL.
FT DOMAIN 103 134 COIL 1A.
FT DOMAIN 135 147 LINKER 1.
FT DOMAIN 148 246 COIL 1B.
FT DOMAIN 247 263 LINKER 12.
FT DOMAIN 264 285 COIL 2A.
FT DOMAIN 286 289 LINKER 2.
FT DOMAIN 290 410 COIL 2B.
FT CARBOHYD 47 47 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CARBOHYD 430 430 O-LINKED (GLCNAC) (BY SIMILARITY).

```


FT	CONFLICT	432		S -> F (IN REF. 2).	
FX	CONFLICT	539		QA -> RR (IN REF. 2).	
FQ	SEQUENCE	848 AA:	95910 MW;	5F251F274D0F13B6 CRC64;	
<hr/>					
Query Match					
Best Local Similarity 3.98; Score 136.5; DB 1; Length 848;					
Matches 157; Conservative 104; Mismatches 270; Indels 249; Gaps 34;					
<hr/>					
Qy	3 SRKSAALDPQSSESTFSSTLDSSAPSNRLRRSGRNILQPSSEKHDKHREGRSEEL	62			
Db	: :				
Qy	53 TRSVAVPLAYSSAMLSAESLDFSSSSLNGSGGDYKLRSNEK-----EQL	103			
Db	: :				
Qy	63 AG---RMKGKDANGHCUREKEQEKGMAIEGLAR-----MERRLQRAT--	104			
Db	: :				
Qy	104 QGLNDREAGYTEKVHYL---EQONKEAEATEIALRQKOAASHAOLGDAYDOEIRELPAITLE	160			
Db	: :				
Qy	105 ---KKQKQ-----LEEDGIPVPVSVRPTAPYHHKSTNAERAKEPVLKTHSKOVE	155			
Db	: :				
Qy	161 MVNIEKAQVLDSDHEED---IHRUKERF-----EEEARLRDDTEAAIRALKRDE	209			
Db	: :				
Qy	156 REAEGVDVVVKMEPATNTII-----EPEDAADAERGAARPPAVNSSYLP-	202			
Db	: :				
Qy	210 ESSMVVELDKKYQSLQDEVAFLLRNHHEEVADLLAQIOASHITVERKDYLTDTISTALK	269			
Db	: :				
Qy	203 -----PWKRLGYACL-----NTYLRNAKPPI-----FSRTCR	231			
Db	: :				
Qy	270 EIRSOLECHSDONHQAEW-FRCRYAKLTEAAEQNKAEIRSAKEEIAEYRQLOSKSTE	328			
Db	: :				
Qy	232 MASIVDRHPQLQ----FEDEPEHLKNPKDSKEPDEL-CHKFVOELGLANARDI--V	283			
Db	: :				
Qy	329 LESVGTGESLERQLSDIERHNHDUSSODTTOOLENELRGTKMEMARHLREYQDLNLV	388			
Db	: :				
Qy	284 KMLCWNKEYGIRELRISSEMFPPASH-----PVHGKYL-----AP----	318			
Db	: :				
Qy	389 KMALDIEIAAYRKL-LEGRETRESTFGSTGLTYHRQPSTVISKTKTKVEAPKLKV	447			
Db	: :				
Qy	319 ---FASVLAAGRVAELGRLHTTHPGQTQGSRKRVESAI RDLEYHDELLSLKL	375			
Db	: :				
Qy	448 OHRFVEEIIET-KVDEKSEMSETITATAELAASAKEEKEBEAPEAKESPVKS	506			
Db	: :				
Qy	376 PEQNDRDAVMIIHMGGFGDKAATLRFKRNYARLSOSCKNRLVLENDVGTVHDDLVP	435			
Db	: :				
Qy	507 PEAKEE-----EAGEKEEEGEQEE-----EEDEG-----	534			
Db	: :				
Qy	436 CBEUNTPMWLYHHNICFPDPAHLREGTLIDSDPKLOERIANTWRKGIKQKMHSPECD	495			
Db	: :				
Qy	535 -----VKSDQA-BEGG-----SEKESSEKDEGEQEEEEGETAE	568			
Db	: :				
Qy	496 GAVTPDRKRHRPVMTLPCPDMDLITEAKDKCAQVFELMRTFKLPGFEKINDMVPYD	555			
Db	: :				
Qy	569 GEGERAEAKEEK-----KTIGRVEEVAVEKEIKVEK-P-EKAKSPMPKS	610			
Db	: :				
Qy	556 RDBENPAPPVPKAPK-----KKKGKRRKTTDEEAEPEDVDTAADDYKDAPEGPK-EVP	609			
Db	: :				
Qy	611 PVEEVKPEAKAGKEQKEEBKEEVKTSKPEKVEKKEPKDVADKKKAESP	670			
Db	: :				
Qy	610 EEERAMGGPNRYWPPLGCSEWL-----RPKKREVKKGVPEVEDEG	652			
Db	: :				
Qy	671 VKEKAV-----EEVITISKSVKVSLEKDTKEERPQOE-----KVKEKAEEG	713			
Db	: :				
<hr/>					
RESULT	6				
ID	CALD_HUMAN	STANDARD;	PRT;	793 AA.	
AC	Q05682; Q13979; Q13978; Q14741; Q14742;				
DC	01-NOV-1997 (Rel. 35, Created)				

DR	MM: 114213; -	
DR	InterPro: IPR000075; Caldesmon.	
DR	Pfam: PF02029; Caldesmon; 1.	
DR	PRINTS: PR01076; CALDESMON.	
KW	Muscle protein; Actin-binding; Calmodulin-binding; Phosphorylation;	
KW	Repeat; Alternative splicing.	
FT	DOMAIN 319 375	3 X 14 AA TANDEM REPEATS OF E-E-K-R-A-
FT		A-E-R-Q-R-I-K.
FT	REPEAT 319 332	1.
FT	REPEAT 333 346	2.
FT	REPEAT 347 360	3.
FT	DOMAIN 26 207	MYOSIN AND CALMODULIN-BINDING (BY
FT		SIMILARITY).
FT	DOMAIN 564 621	TROPOMYOSIN-BINDING (POTENTIAL).
FT	DOMAIN 664 674	TROPOMYOSIN-BINDING (POTENTIAL).
FT	DOMAIN 653 686	STRONG ACTIN-BINDING (BY SIMILARITY).
FT	DOMAIN 716 722	CALMODULIN-BINDING (BY SIMILARITY).
FT	DOMAIN 768 793	WEAK ACTIN-BINDING (BY SIMILARITY).
FT	DOMAIN 39 46	POLY-ARG.
FT	DOMAIN 81 86	POLY-THR.
FT	DOMAIN 189 196	POLY-GLU.
FT	DOMAIN 376 379	POLY-GLU.
FT	DOMAIN 540 543	POLY-ARG.
FT	DOMAIN 580 583	POLY-GLU.
FT	DOMAIN 597 600	POLY-GLU.
FT	MOD_RES 724 724	PHOSPHORYLATION (BY CDC2) (BY
FT		SIMILARITY).
FT	MOD_RES 730 730	PHOSPHORYLATION (BY CDC2) (BY
FT		SIMILARITY).
FT	MOD_RES 753 753	PHOSPHORYLATION (BY CDC2) (BY
FT		SIMILARITY).
FT	MOD_RES 759 759	PHOSPHORYLATION (BY CDC2) (BY
FT		SIMILARITY).
FT	MOD_RES 789 789	PHOSPHORYLATION (BY CDC2) (BY
FT		SIMILARITY).
FT	VARSPLIC 1 24	MODERRRRLRQKREMRLEABER -> MLGSGSHGRRSL
FT		AALSO (IN ISOFORM HELA L-CAD I AND
FT		ISOFORM HELA L-CAD II).
FT	VARSPLIC 208 436	MISSING (IN ISOFORM HELA L-CAD I AND
FT		ISOFORM WI-38 L-CAD I).
FT	VARSPLIC 208 462	MISSING (IN ISOFORM HELA L-CAD II AND
FT		ISOFORM WI-38 L-CAD II/I-L-CAD).
FT	CONFLICT 530 530	V -> M (IN REF. 1).
FT	SEQUENCE 793 AA; 93250 MW; 2A0DC63D16D6B5F CRC64;	
SQ		
Query Match 3.9%; Score 135; DB 1; Length 793;		
Best Local Similarity 20.5%; Pred. No. 0.53;		
Matches 143; Conservative 105; Mismatches 271; Indels 180; Gaps 28;		
QY	34 RNLRRSGNLPQSEKDRHDKRSGEELAGRMCKDANGHCLREGKQEE	----- 83
DB	8 RELURQKREMRLEABRIAYQRNDDDEFAARERRRRARRQERLRQKQBEESLGQVTDQVE	67
QY	84 -----EGVKMA--IEGLARME-----RRLQRAATKQKK-----QLEE 113	
DB	68 VNAQNSVPDEAKTTTNTQVCDDEAAFLRLARERRRQRLQALERKEFDPTITD	127
QY	114 DGIPVS-----VSRFPATPYHHSKTAER-----EAKEPVLKTHSKDVERAEIGVDV 165	
DB	128 ASLSLPSRRMQNDTAENNETTEKSESQERYEIEETETVTKSYQKNDRDAENKKED	187
QY	166 VKMEPATNLIPEDDAADAERGAARPPAVNSSYLPPLPKWGLRGVACLYTLRNAPPIF	225
DB	188 KEKE-----EEEEPKKG-----SIGENQEVVWVEEKTTSQETVVMYSLKNGOI 233	
QY	226 SSRTCRM-----ASIVDRHPLOFED-----EPDHILKNPKDSKEPQDELIG 267	
DB	234 SSEPQKEEREQSGDSIEHSHEKMEBEDKRAEAEARLEAEFERIKAEQDKKIADEPA 293	
QY	268 HKFVQELGLANARDIVKMLCWNKEYGIRFLRLSLSEMPFPASHPVHGYKIAPASEVLAE 327	
DB	294 RIAEAKAAQERE-----RRRE-----EREMREE 320	

QY	328 GFVAALGHLHTTTHPGQFTQLGSPRKEVVESSAIRDLEYHDELLSLLLKLPQONRDAMII	387
DB	321 EKRAAEERQRIKE---EEKRAAEERQRIKEEKRAAEERQ-----RIKEEKRAAEERQ	371
QY	388 HMGQFGDKAAATLIERKRYARLSOSKNRLVLENDVGVTHDLLPVCBELNIPMVDY 447	
DB	372 RARAEERKE-ARVEQKRN-KOLEE--KRAMQETKIKKEVE-----QKIEGKWNEK	421
QY	448 HHNHCIFDPAHLR-----EGT-LDISDPKLOERIANTWKKGIKOK--MHYSEPCDGAVT	499
DB	422 KAOEDKLOTVAVLKQGEKEGTQVQAKREKLQED-KPTFKKEIKDEKIKKKEPEEYKS	480
QY	500 PRDRKRHRPRVMTLPPCPDMDLMEAKDKQAVFELM---RTFKLPGEFEXINDMVPYDR	556
DB	481 FMDRKKGFTEV-----KSONGEDMTKHLKHTENTFSRPGGRASVD-----	520
QY	557 DQNRPAIPVAPKPKKKKGKRR--TTDEE-----AAEPEVDTAADDVKDAPEGPKVEPE	610
DB	521 TREAGAPQVEAGKRLLELRRRRGETESEFEKLOKQOQEALELELKKRERKKVLE	580
QY	611 EERAMGGPNRVVWPLGCRWLKPKKREVKGKVPPEE 649	
DB	581 EBEQRKQ-----EADRLPEEEERKRLKEEIE 609	
RESULT 7		
ID	CENE HUMAN STANDARD; PRT: 2663 AA.	
AC	Q02224;	
DT	01-JUL-1993 (Rel. 26, Created)	
DT	01-JUL-1993 (Rel. 26, Last sequence update)	
DT	30-MAY-2000 (Rel. 39, Last annotation update)	
DE	CENTROMERIC PROTEIN E (CENP-E PROTEIN).	
GN	CENPE.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID:9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE-93024922; PubMed-1406971;	
RA	Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;	
RT	"CENP-E is a putative kinetochore motor that accumulates just before	
RT	mitosis";	
RL	Nature 359:536-539(1992).	
RN	[2]	
RP	CHARACTERIZATION.	
RX	MEDLINE-95196755; PubMed-7889940;	
RA	Theriot D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;	
RT	"Mitotic HeLa cells contain a CENP-E-associated minus end-directed	
RT	microtubule motor";	
RL	EMBO J. 14:918-926(1995).	
RN	[3]	
RP	CHARACTERIZATION.	
RX	MEDLINE-98437347; PubMed-9763420;	
RA	Chan G.K.T., Schaar B.T., Yen T.J.;	
RT	"Characterization of the kinetochore binding domain of CENP-E reveals	
RT	interactions with the kinetochore proteins CENP-F and hBUBR1.";	
RL	J. Cell Biol. 143:49-63(1998).	
CC	-1- KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE	
CC	OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT	
CC	AND/OR SPINDLE ELONGATION.	
CC	-1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.	
CC	-1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING	
CC	CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS	
CC	QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.	
CC	-1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.	
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: Z15005; CAA78727.1; -
 PIR: S28261; S28261.
 HSSP: P03069; 1211.
 MIM: 117143; -
 InterPro: IPR001752; kinesin.
 Pfam: PF00225; kinesin; 1.
 PRINTS: PR00380; KINESINHEAVY.
 SMART: SM00129; KISC; 1.
 PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 PROSITE: PS50067; KINESIN_MOTOR_DOMAIN2; 1.
 Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
 Cell cycle; Centromere.
 FT DOMAIN 1 335 MECHANOCHEMICAL (MOTOR).
 FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
 FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
 FT NP_BIND 86 93 ATP (BY SIMILARITY).
 SQ SEQUENCE 2663 AA: 312087 MW: 45713880C8C8CB8 CRC64;

Query Match 3.8%; Score 134; DB 1; Length 2663;
 Best Local Similarity 17.9%; Pred. No. 2.8;
 Matches 143; Conservative 136; Mismatches 280; Indels 252; Gaps 32;

QY 16 SESSTFSLDS-----SAPSPARNLRSGRNI----- 43
 DB 467 SESOVFNTLDTLSEHWNPATKLLNQENIESLNLRADYDNLVDYDQLRTEKEEMEL 526
 QY 44 -----LPSSEKDRD-----HEKRSGBELAGRMGMKGANGHC 75
 DB 527 KLKKNLDLDEFEALERTKKDQEMQLIHEISNLKLVKHEVYNQDLENELSK---VEL 583
 QY 76 LRGGKQ-----EEGVKM-----AIEGLARMERLQRA-----TKRKKO 110
 DB 584 LREKEDIKRLQIYDSQLENIKMDLSYSLESIEDPKQMKOTLFDAETVALDAKRESAF 643
 QY 111 LEEDGIPVPSVWSRFPATPVHKKSTNAEREAKEPVLKTHSKDVEREALGVDDVVVKPEP 170
 DB 644 LRSENLEKMKELATT-YKOMENDIQLYQSOLEAKKMKQVDEKLSQAFNEITKL-- 700
 QY 171 AATNIE---PEDA-----QDAAEARGAAPPAVNSSYL-PLPWKGR 208
 DB 701 --FSLIDGKVPKDLLCNLEGGKITDLQELNKEVEENALREEVILLSELKSLP----- 753
 QY 209 GYACLNTYLRNAKPPIFSSTRTCMASIVDHRHPLQFEDPEPHLLKNKPKSKPEQDELGH 268
 DB 754 -----SEVERLRKEIQDKSEELHIITSEKDKL---ESEVHV 786
 QY 269 K-----FVQELG-----LANARDIVK-----MLCWNKEYGIR 295
 DB 787 KESRVGLIEETGKTDDLIATTSQNTKSTQDFQNFKTLHMDFEOKYKMWLEENRMQOE 846
 QY 296 FLRLSEMFPFASHPHVGHYKLAFFASEVLAEEAGRAAELGHLRLTHPGQFTQLGSPPKEV 355
 DB 847 IVNLSKEAQKFD-----SLGALKTELKSYK-----TQELQEKTR 881
 QY 356 VESAIRDLEYHDELLSLKLPEQO-NRDAMVIHMGQFGDKAAALTERPKRNYARLSQSC 414
 DB 882 VOERLNEMLQKLEQNRSPLQTVREKXTLITEKLOQTLEEVKTLTQEKDKDLQKQES- 940
 QY 415 KNRVLNDVDGVTVHDL---LPVCEEL-NIPMVLVDYHHNI-----CFDPAHL 459
 DB 941 ---LQIERQLKSDIHDYNNMIDTQEQURNALSLKQHOETINTLKSKTSEVSRNLHM 997
 QY 460 RECTLDISPKLQBIERTANTWRKGIKQKMHYSEPCDGAVTTPDRKRKRPVMTLPCCPD 519
 DB 998 EENTGETKD-EFQKQWGVGDKKQDLKAKNTQTLAD-VKDNIEITQQRKIFSLQEKNE 1054
 QY 520 MDLMIEA--KDEQAVFELMRTFK--LPGFEKINDMVPYDRDDENRAPPPVAPKK---- 571

DB 1055 LQOMLESVIAKEQLKTLKENIEMTIENQBELRL-----GDELKKQOEIVAEKNHAI 1109
 QY 572 KGGKRKRRTTDEAAPEEVDTAADDVKDAPEGPKEVPEERAMGPGPNRYVWPLGCEW 631
 DB 1110 KKEGELSRCTDLRAEVEEKLKESQQLQEQKQQLLNVOFENSEMOKKINEI----- 1160
 QY 632 LKPKKREVKKGKVP-EEVEDE 651
 DB 1161 -ENLKNELKNKELTLEHMETE 1180
 RESULT 8
 ZDSL_YEAST
 ID ZDSL_YEAST STANDARD; PRT: 915 AA.
 AC P50111; P41373;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE ZDSL PROTEIN (NRC1 PROTEIN) (RT2GS1).
 GN ZDSL OR NRC1 OR CES1 OR CKM1 OR OSS1 OR STM2 OR HST1 OR YMR273C OR
 GN YMR156.15C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=964113277; PubMed=8816439;
 RA Bi E., Pringle J.R.;
 RT "ZDSL and ZDS2, genes whose products may regulate Cdc42p in
 RT Saccharomyces cerevisiae";
 RL Mol. Cell. Biol. 16:5264-5275(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Schwer B., Jiu Y., Shuman S.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96249381; PubMed=8647431;
 RA Ma X.J., Lu Q., Grunstein M.;
 RT "A search for proteins that interact genetically with histone H3 and
 RT H4 amino termini uncovers novel regulators of the Swi1 kinase in
 RT Saccharomyces cerevisiae";
 RL Genes Dev. 10:1327-1340(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COP161U7;
 RA Jia Y., Li J.-P., Butow R.A.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Linder P.;
 RL Submitted (AUG-1994) to the SWISS-PROT data bank.
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RAY-3A;
 RX MEDLINE=97075905; PubMed=8918228;
 RA Tsuchiya E., Matsuzaki G., Kurano K., Fukuchi T., Tsukao A.,
 RA Miyakawa T.;
 RT "The Saccharomyces cerevisiae Ssn1 gene is involved in the tolerance
 RT to high concentration of Ca2+ with the participation of
 RT HSP1/NRC1/BPR1";
 RL Gene 176:35-38(1996).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Lye G., Churcher C.M., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: NOT KNOWN; MULTICOPY SUPPRESSOR OF A VARIETY OF DEFECTS.
 CC SEEMS TO INTERACTS WITH CDC42. MAY PLAY AN IMPORTANT ROLE IN CELL
 CC CYCLE PROGRESSION.
 CC -!- MISCELLANEOUS: ZDS MEANS 'ZILLION DIFFERENT SCREENS' AS BOTH ZDSL

```
CC AND ZDS2 HAVE BEEN FOUND BY A WIDE VARIETY OF GENETIC SCREENS.
CC -!- SIMILARITY: STRONG, TO YEAST ZDS2/MCS1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L42821; AAB38423.1; -
CC EMBL: U32580; AAA74626.1; -
CC EMBL: U63849; AAB49281.1; -
CC EMBL: U63682; AAB08617.1; -
CC EMBL: D50276; BAA08819.1; -
CC EMBL: Z49260; CAA89256.1; -
CC SGD: S0004886; ZDS1.
CC
CC DOMAIN 716 719 POLY-GLU.
CC FT DOMAIN 907 914 POLY-GLN.
CC FT CONFLICT 84 84 S -> P (IN REF. 5).
CC FT CONFLICT 393 393 D -> E (IN REF. 2).
CC FT CONFLICT 413 413 S -> C (IN REF. 2).
CC
CC SEQUENCE 915 AA; 103358 MW; E5F4C2117DBE1210 CRC64;
CC
Query Match 3.8%; Score 132.5; DB 1; Length 915;
Best Local Similarity 18.18; Pred No. 0.9;
Matches 130; Conservative 91; Mismatches 216; Indels 281; Gaps 30;
CC
QY 12 DTPOSE-----SSTFSTLD-----SSAPS-----PAR-NLRSGRNILQPSSE 49
DB 317 DQPOEGHYDEGDIGFSTSQANTLLDGEFASNMPINNTMTWPERSLRRSFRNTYRISQ 376
QY 50 KDRHEKSGELACRMGMKDANGHCLREGKEQEGVKMALEGRLARRLORATKROKK 109
DB 377 EOEKEVOSVDEM-----NDDEERLUKT-KNTIKVEIDPHSPFFQQD 419
QY 110 QLEEDGIPVSVSRFPPTAPYHHKSTNAE-----BREAPEVLKTHSKDVERAEI-- 160
DB 420 E-DSNNMSPSGISGDFQDIYHYRQSGGEQEMGIEKEAEVVKVRNDTVEQDLLE 478
QY 161 GVDVVVKMEPATNIIPEDAQDAERCAARPPAVNSYLLPWPWKGRIGVACLTLYRNA 220
DB 479 GTDMVK--PSAT-----DDNKETKRH-----RRNGTWLANKM--- 511
QY 221 KPPIFSRTCRMASTVHRHPLQFDEPEHLKKNPKPKPEDELGHKFVQELGLANAR 280
DB 512 -----SREDDNEENGDDENEENVD-----QRMELDNK 541
QY 281 DIVKMLCHNEKYGIFRLSLSMFPFASHPVHYKLPAPFASEVLAEGRAVAELCHRLTT 340
DB 542 -----KHYISLFGNGEKTEVSNNKEEMNSSTSTATSQTOKTEKTFANLFRKRPH 591
QY 341 HPCQPTQLGSPKRVESAIROLEYHDELLSLLKLPEQONRDVMI-IHMGQGPDKAAAT 399
DB 592 H--KHDASSSSSSPSS-----PSIPNDVAHVVRVRSKSLKNGSG- 631
QY 400 LERFRKNRYLSQSCKNRLVLENDVGWTVHDLPLVCEELNIPVVL-----DYHHNICF 454
DB 632 -----REPVEIVLRNRPFRHRH----- 650
QY 455 DPAHLREGTLDIS-----DPKLOERIANTKRKGIKQRMHYSEPCDGAVTDRRKRHR 509
DB 651 -----HSRHGSKISVTKLDSQPOQI----- 673
QY 510 VMTLPPCPDMDLMIKAEKQAVFELMRTEKLPGE-----KINDMVPYDRDNRPPVP 566
DB 674 -----PLOQLBGAIEIKKEESDSE-----SLQLOQPAVSVSTKSNRDRBEAEA--- 720
QY 567 KAPKKKKGKGR-----KRTTDEAAE-----PEEVDTAADDVKDAP--- 602
DB 721 -----KKKNKRSNTTEISNOOHKSHVOKENTDEQAKLQAPAEQVQTSVVPVQASAPVN 776
```

```
QY 603 -----EGFKVPEERAMGGYNNVYVPLGCEEWLKPCKREVKKGVPEVEDEGEF 654
DB 777 SAPVQTSFAVEASQTOQAPAP-----PLKHTSLTPPKRKUTFADVKKPKDPKNSPVQF 828
```

RESULT 9

```
PLECT_CRIGR STANDARD; PRT: 4473 AA.
AC Q9J155;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PLECTIN 1 (PLFN) (PCN) (300-KDA INTERMEDIATE FILAMENT-ASSOCIATED
DE PROTEIN) (IFAP300) (FRAGMENT).
GN PLECT.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=20334248; PubMed=10873583;
RA Clubb B.H., Chou Y.-H., Herrmann H., Svitek T.M., Borisy G.G.,
RA Goldman R.D.;
RT "The 300-kDa intermediate filament-associated protein (IFAP300) is a
RL Biochem. Biophys. Res. Commun. 273:183-187(2000).
RN [2]
PHOSPHORYLATION.
RX MEDLINE=96215219; PubMed=8626512;
RA Malecz N., Folsner R., Stadler C., Wiche G.;
RT "Identification of plectin as a substrate of p34cdc2 kinase and
RT J. Biol. Chem. 271:8203-8208(1996).
CC -!- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
CC MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSOMES OR
CC HEMIDESMOSOMES. MAY BE INVOLVED NOT ONLY IN THE CROSSLINKING AND
CC STABILIZATION OF CYTOSKELETAL INTERMEDIATE FILAMENTS NETWORK, BUT
CC ALSO IN THE REGULATION OF THEIR DYNAMICS.
CC -!- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
CC -!- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH
CC VIMENTIN, DESMIN, GAP, CYTOKERATIN, LAMIN B; WHEREAS BOTH THE N-
CC AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.
CC -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM
CC INTERMEDIATE FILAMENTS DURING MITOSIS.
CC -!- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY DOMAINS (CH).
CC -!- SIMILARITY: CONTAINS 32 PLECTIN REPEATS.
CC -!- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
CC -!- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF260753; AAF70372.1; -
CC InterPro: IPR001101; Plectin_repeat.
CC InterPro: IPR001589; Actinin_act_bind.
CC InterPro: IPR001715; Calponin_hom.
CC InterPro: IPR002017; Spectrin.
CC Pfam: PF00307; CH; 2.
CC SMART: SM00681; Plectin_repeat; 20.
CC Pfam: PF00681; Plectin_repeat; 20.
CC SMART: SM00033; CH; 2.
CC SMART: SM00250; PLEC; 32.
CC SMART: SM00150; SPEC; 4.
CC PROSITE: PS00019; ACTININ_1; PARTIAL.
CC PROSITE: PS00020; ACTININ_2; FALSE_NEG.
```


CC -!- FUNCTION: MAY REPRESENT A CHROMATIN-ASSOCIATED ACETYLTRANSFERASE.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- DISEASE: PARTICIPATES IN A T(8;16)(P11;P13) CHROMOSOMAL
 CC TRANSLOCATION THAT PRODUCES A MOZ-CBP CHIMAERA OBSERVED IN THE
 CC M4/M5 SUBTYPE OF ACUTE MYELOID LEUKEMIA.
 CC -!- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
 CC -!- SIMILARITY: BELONGS TO THE MYST (SAS/MOZ) FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U47742; AAC50662.1; --
 CC HSSP; Q50631; 4GB0.
 CC MIM; 601408; --
 CC InterPro; IPR001386; Linker_histone.
 CC InterPro; IPR002717; MOZ_SAS.
 CC InterPro; IPR001965; PHD.
 CC Pfam; PF01853; MOZ_SAS; 1.
 CC Pfam; PF00628; PHD; 2.
 CC SMART; SM00526; HL5; 1.
 CC SMART; SM00249; PHD; 2.
 CC Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;
 KW Nuclear protein.
 FT 2N_FING 206 256 PHD-TYPE 1.
 FT 2N_FING 259 313 PHD-TYPE 2.
 FT DOMAIN 371 379 POLY-SER.
 FT 2N_FING 538 560 C2HC-TYPE.
 FT DOMAIN 788 801 POLY-GLU.
 FT DOMAIN 989 995 POLY-GLU.
 FT DOMAIN 1019 1026 POLY-ARG.
 FT DOMAIN 1069 1078 POLY-GLU.
 FT DOMAIN 1147 1150 POLY-LYS.
 FT DOMAIN 1221 1242 GLU-RICH.
 FT DOMAIN 1267 1302 GLU-RICH.
 FT DOMAIN 1411 1414 POLY-GLU.
 FT DOMAIN 1593 1597 POLY-SER.
 FT DOMAIN 1643 1704 GLN/PRO-RICH.
 FT DOMAIN 1897 1977 MET-RICH.
 FT SITE 1546 1547
 FT
 FT BREAKPOINT FOR TRANSLOCATION TO FORM
 FT MOZ-CBP.
 FT
 FT SEQUENCE 2004 AA: 225054 MW: 9FBBAC3792854BA CRC64;
 FT
 Query Match 3.7%; Score 128.5; DB 1; Length 2004;
 Best Local Similarity 20.1%; Pred. No. 4.3;
 Matches 136; Conservative 89; Mismatches 260; Indels 193; Gaps 32;
 QY 47 SSEKDRDKEKSGEELAGR-----MMGKDANGHCLREGKEQEGVKMAIEGLARMERRL 100
 Db 787 SEEEEEAEAGEENEPQCOERELEISVGKSVS-----HENKED-----SYSVSEKKPEVMA 839
 QY 101 QRATRKQKQ-LEEDGIPVPSVRPTAPYHHKSTNAEER-----EAKPEVLKTHSKDVE 155
 Db 840 PVSSTRLSKQVLPBDSLANSQPSR--RGRWGRKNRKTQERFGDKSKLLLEETSSAPQE 897
 QY 156 REAEIGVDVVKWEKPAANTNIEPEDAQDAERGAARPAVANSYLP-----PWKGR 207
 Db 898 QYGECE-----EKSEATQOYTESBOLVASEE---QPSQDGKPDLPKRRRLSEGVEPWRGQ 950
 QY 208 LGYACLTNLYRNAKPPIFSSRRCTMASIVDHRHPLQFDEPEHLLKNPKDKSKPEQDELG 267
 Db 951 L-----KKSP--EALKCRUTE-GSERLPRRYSEGDAVLRGFSSESEEEPEPS 996
 QY 268 HRFVQELGLANARDIVKMLCWNEKYGIRFLRSLSEMFPFASHVGHYKLPAPFASEVLAEA 327
 Db 997 PR-----SSSPILTKPTLKRK-KPF-----LHRR 1020
 QY 328 GRVAAELGHRLTHPGQFTQLGSPKRVESAIRDLEHYDELISL---LKLPEQONRDV 384

Db 1021 RRVRRKKHINSV-----VTETISSETTEVLDEPFDSDSERPMRPLEPTFEIDEEEEE-- 1074
 QY 385 MIIHMGOGFGDKAATLERFKRNY-----ARLSQSKNRLVLEDDVGVTVHDLPLV 435
 Db 1075 -----DEN-----ELFPREYFRLLSQDVLRQSSSKRKSDEEEDSDADDPTPI 1121
 QY 436 CEELNIPVLDYHHNICFD---PAHLREGTLDISDPKLOERIANTW-KRKGIOKKMHYS 491
 Db 1122 LKPVSLLRKRDVKNSPLEPDTSTPLKKGW-----PKGSKRKP1HMKKRPGRKPGPKLS 1176
 QY 492 -----EPCDGAVTPRDRKRPRVM-----TLPCPPDMDLMIEAKDEQAVF 534
 Db 1177 REIMPVSTQACVIEPI--VSPKAGRK--PKIQESEETVEP-KEDMPUPPEERKEEEMOA 1231
 QY 535 ELMRTFKLPGEKINDMVP--YDRDENRPAAPPVAPKPKKGGKKRRTTDE----- 583
 Db 1232 EAAEAEEGEEEDAAASSEVPAAADSSNPETETKEPEVEEEEEKPRVSEBQROSEEQO 1291
 QY 584 --EAAEPEEVDTA-----DDVKAPEGPKVEPEERAMGGPYNRVYVPLGCEEWLK 633
 Db 1292 ELEPEPEEEEDAAAEATAQNDHDADEDDGHLE-----S 1326
 QY 634 PKKREVKKGVPEEVEDE 651
 Db 1327 TKKKELEEOPTREDVKEE 1344
 RESULT 12
 NFH_HUMAN
 ID NFH_HUMAN STANDARD; PRT; 1020 AA.
 AC P12036;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
 GN NFH OR NFH.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88328981; PubMed=3138108;
 RA Lees J.F., Sheldman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;
 RT "The structure and organization of the human heavy neurofilament
 subunit (NF-H) and the gene encoding it."
 RL EMBO J. 7:1947-1955(1988).
 CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
 PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 OF AXONAL CALIBER.
 CC -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X15306; CAA33366.1; -.


```
DR EMBL; X15307; CAA33366.1; JOINED.
DR EMBL; X15308; CAA33366.1; JOINED.
DR EMBL; X15309; CAA33366.1; JOINED.
DR PIR; S00979; QFHUH.
DR MIN; 162230; -.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation.
FT DOMAIN 1 100 HEAD.
FT DOMAIN 101 413 ROD.
FT DOMAIN 414 1020 TAIL.
FT DOMAIN 101 132 COIL 1A.
FT DOMAIN 133 145 LINKER 1.
FT DOMAIN 146 244 COIL 1B.
FT DOMAIN 245 266 LINKER 12.
FT DOMAIN 267 288 COIL 2A.
FT DOMAIN 289 292 LINKER 2.
FT DOMAIN 293 413 COIL 2B.
SQ SEQUENCE 1020 AA; 111780 MW; 1177C9DCB3DCF1D4 CRC64;

Query Match 3.7%; Score 128; DB 1; Length 1020;
Best Local Similarity 20.9%; Pred. No. 1.9;
Matches 130; Conservative 84; Mismatches 243; Indels 164; Gaps 30;

QY 45 QPSSEKDRHEKRSGEELAGRMGKDGANGCHLRECKEKEGKVAIEGLARMERLQRA 104
DB 465 QVTEETVEEKEAKEE-----EKKEEGEGEEAEAG-GEEETKSPAE 507

QY 105 KRQKQLEEDGIPVPSVVSFRPTAPYHHKSTNAEREKEPV-LKTHSKDVEREAET 163
DB 508 EAASPEKE-----AKSPVKEAKSPAEEKS--PEKEEAKSPAEEKSPKEK 555

QY 164 DVVKMEPATNIIEPEDQAADAEAGA--ARPPAVNSYLLPLPKWGLRGYACLA 220
DB 556 EPAKSPPEAKS-PEKEEAKSPAEEKSPEKAKSPAEEKSPAEEKSP-----EKA 604

QY 221 KPPISSRTCRMASIVDHRHPLQFDEPEHILKNPKDKSKPEQDPLGHKFVOELGLANAR 280
DB 605 KSPVKE-----EAKSPAEEKSPAEEKSPAEEKS--PEKASPTKEEAKS--PEKASPEK 657

QY 281 DIVKMLCWNEKGIKRIPLISSEMPFPASHVGHYKLP--FASEVLADAGRAVAELG 338
DB 658 EEAK-----SPEK-----AKSPVKAEEKSPAEEKSPAEEKS-----688

QY 339 TTHPGQFTOLGSPKRVESATRDLEYHDELLSLKLPEONRDVAMIHMGGQFGDKAA 398
DB 689 -----KSPKE--AKSPVKE-----EAKSPEKAK-----SPVKEEAK 717

QY 399 TLERFK---RNYARLSQCKNRLVLENDVGTWVTHDLLPVGCEELNIPMVDYHHNICFD 455
DB 718 SPEKAKSPVKEEAKTPEKAKS-----PVKEEAKSP-----E 748

QY 456 PAHLRE--GTLDISPKLOERTANTWKRKGIQKMHYSEPCDGYTPDRDKRHRPVMTL 513
DB 749 KAKSPEKATLDVKSPEAKTPEAKSPAKESPADKFPPEKAKSPVKEEVKSPPEKAK-----800

QY 514 PCPPDMDLMIEAKDEQAVFELMRTFKLPGEKINDMVPYDRDDENRPAPEVPAKPKKK 573
DB 801 SPLKADAKAPEKIPKKEEVKSPVKEEKPQEVKVE--PPKAAEKAAPATP--KTEEKK 857

QY 574 GGRKRTTDEEAAP---BEVDTAADVVDKADP--EGPKVEPEERAMGPGYNYWPLGCE 629
DB 858 DSKKEAPKKEAPKPKVEBKKEFAVEKPKESKVEAKKEAEADKKKVPPTPEKAPAKVEK 917

QY 630 EWLKPK-KREVKKGKVPERVE 649
DB 918 EDAPKPEKTEVAK-KEPDDAK 937

RESULT 13
NEST_RAT
```

```
ID NEST_RAT STANDARD; PRT; 1805 AA.
AC P21263;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NESTIN.
GN NES.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN 111
RP SEQUENCE FROM N.A.
RX MEDLINE=90150286; PubMed=1689217;
RA Lendahl U., Zimmerman L.B., McKay R.D.G.;
RT "CNS stem cells express a new class of intermediate filament
protein.";
RL Cell 60:585-595(1990).
CC -!- TISSUE SPECIFICITY: CNS STEM CELLS.
CC -!- DEVELOPMENTAL STAGE: UPON TERMINAL NEURAL DIFFERENTIATION, NESTIN
CC IS DOWN-REGULATED AND REPLACED BY NEUROFILAMENTS.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M34384; AAA41685.1; -.
CC PIR; A34736; A34736.
CC InterPro; IPR001664; IF.
CC Pfam; PF00038; filament; 2.
CC PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurone.
FT DOMAIN 1 7 HEAD.
FT DOMAIN 8 314 ROD.
FT DOMAIN 315 1805 TAIL.
FT DOMAIN 8 43 COIL 1A.
FT DOMAIN 44 55 LINKER 1.
FT DOMAIN 56 151 COIL 1B.
FT DOMAIN 152 174 LINKER 12.
FT DOMAIN 175 193 COIL 2A.
FT DOMAIN 194 196 LINKER 2.
FT DOMAIN 197 314 COIL 2B.
SQ SEQUENCE 1805 AA; 198744 MW; B40EE14717E0998D CRC64;

Query Match 3.7%; Score 128; DB 1; Length 1805;
Best Local Similarity 18.8%; Pred. No. 4;
Matches 167; Conservative 102; Mismatches 287; Indels 334; Gaps 41;

QY 13 TPQSESTFSSTLDSSAPARNLRKSRNIIQSPSEKDRDHEKRSGE---LAGRMVG 69
DB 685 SPEEDEACRPLQKNEQEPGLGYEEAEG-ILERLIEKESQESILRSPPEEDQEA 743

QY 70 -----DANGHCLRE--GKEQEGCVKMAIEGLARMERLQRAKROKLEEDG---115
DB 744 ENQPLGYEEAEQMLPRLIEKESQESLSKSPEN-ORIGKPLERENOKSLRYLEENQ 802

QY 116 IPVPSVVSFRPTAPYHHKSTNAEREKEPVLTSHKSDVEREAIEGV-----DDVVK 167
DB 803 VPLESNRQR---PLRSLEVEEEQRIKVEKVSODSLGSLAEENVQPLRYLEEDDCIN 858

QY 168 M-----EPAATNIIEPEDQAADAEAGAARPPAVNSYLLPLPWKGLRGYAC 212
DB 859 KSLLEDKTHKSLGSLDRNGDSIIIPQSESTQV---SLRPEEEDQRI-----903

QY 213 LNTYLRAKAPPIFSRTRCMASIV-----DHRHPLQFEDPEPEHLKN 254
DB 904 VNHLEKESQE---FSRSSEEEQVMEKSELEGENHIESLSVKEKEDQMVESQLESQSGKS 961
```


QY 255 KPDKSKP-----QDELGHKFOE----- 273
 Db 962 LEDESQTFGLKPKENAESLRSAGQDEQKLEQETQOQLTRAVGNEQMAVSPPEKVDPE 1021
 QY 274 -----LGLANARDIVKMLCWNEKYGIRFLR--LSSEMF----- 304
 Db 1022 LKPLGNDQEIARSLGKNEQESIVSL-----KEGIETVKSLETEIIEPLETAEDEDLERKK 1077
 QY 305 -----PFASHPVHGKLPAPFASEVLAEGRAVAAELGHLRTHHPGQFTGLGSPRKVEVES 358
 Db 1078 SIDQPELWSTEVARETPEPEPPGSLGSV--DENRETLTSLKESQELSLGKKNVET 1136
 QY 359 AIRDL-----EYHDELSLLK--LPEQNR-----DAV----- 384
 Db 1137 RVDSQOCLQVEGLQEQHESLREVKOELPSSGNOQRWEDVVEGKAVGOEAPLATTGV 1196
 QY 385 -----MIIHGGQFGDKATLRFKKNYARLSOSCKNRVLNDDVY--WTVHDLPLVCE 437
 Db 1197 GTEDKAEHLRGGGEEAAEG-----ELQDIVGEAWSLGSSEP--K 1238
 QY 438 ELNIPM-VLDYHHNHCDFPAHLRGCTLDI-----SDPKLOER-----TANTWK 480
 Db 1239 EQRVPAELD-----NLEGALEVPVAQSMPEVTERDEDRQAAGEQDSIEVTLG 1287
 QY 481 RKGIKQMHYSEPCDGAVTPRDRKRHRPRVMTLPPCPPDMDLM-----IEAKDKE----- 530
 Db 1288 LEAARTGLEQEVGLDPR-----HFAREEATPSSLGSESVKAKTAQGLEGFGKEPKEA 1343
 QY 531 -----QAVFELMRT-----FKLPQFEKINDMVPYDR-----DDENRPAPPVKAP--KKK 572
 Db 1344 GALDSGTLELPKRTSSPALECOGHESESEMEGWEERASLETSDHEGSDAPQPRPETED 1403
 QY 573 KGKRRKT-----TD-----EAAPEPEVDTAADVDKADPB- 603
 Db 1404 EGAQAALTAPGPKLLEPCSPILPILDAHELQPAEGIOAGWQPEAGSALERVEVEPEF 1463
 QY 604 GKPEVPEERAMGPNRYNWPVLGCEEWLKKPKREYKKGKVPVEVEDEGE 653
 Db 1464 GLGEIPE-----GLQDW-----EEGRESEADDLGE 1489

RESULT 14

NCRL_MOUSE STANDARD; PRT: 2453 AA.
 AC Q60974; Q60812;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NUCLEAR RECEPTOR CO-REPRESSOR 1 (N-COR) (N-COR) (RETINOID X RECEPTOR INTERACTING PROTEIN 13) (RIP13).
 GN NCOR1 OR RXRIP13.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
 RC TISSUE=Pituitary;
 RX MEDLINE=96008539; PubMed=7566114;
 RA Hoerlein A.J., Naeae A.M., Heinzel T., Torchia J., Gloss B.,
 RA Kurokawa R., Ryan A., Kamei Y., Soederstrom M., Glass C.K.,
 RA Rosenfeld M.G.;
 RT "Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor.";
 RL Nature 377:397-404(1995).
 RN [2]
 RP SEQUENCE OF 1792-2453 FROM N.A. (ISOFORM LONG).
 RC TISSUE=Liver;
 RX MEDLINE=95280959; PubMed=7760852;
 RA Seol W., Choi H.S., Moore D.D.;
 RT "Isolation of proteins that interact specifically with the retinoid X receptor: two novel orphan receptors.";
 RT

RL Mol. Endocrinol. 9:72-85(1995).
 CC -!- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
 CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
 CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
 CC -!- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
 CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
 CC WITH THE THYROID (TR) AND THE RETINOID ACID RECEPTORS (RAR) IN THE
 CC ABSENCE OF LIGAND.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
 CC -!- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
 CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2
 CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-
 CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION
 CC DOMAINS (ID1 AND ID2).
 CC -!- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED
 CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND
 CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES
 CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR
 CC SPECIFICITY.
 CC -!- SIMILARITY: CONTAINS 1 SANT DOMAIN (COMPOSED OF SANT-A-SANT-B).
 CC -!- SIMILARITY: CONTAINS 2 CORNR BOX.
 CC -!- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
 CC FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL: U35312; AAC17125.1; -;
 DR EMBL; U35312; AAC17125.1; -;
 DR MCD; MGI:1349717; Ncor1.
 DR InterPro; IPR001005; MYB_DNA_bind.
 DR Pfam; PF00249; myb_DNA-binding; 2.
 DR SMART; SM00395; SANT; 2.
 DR PROSITE; PS50090; MYB_3; 1.
 DR Nuclear protein; Transcription regulation; DNA-binding; Repressor;
 DR Coiled coil; Alternative splicing.
 KW DOMAIN 174 216 COILED COIL (POTENTIAL).
 FT DOMAIN 254 312 INTERACTION WITH SIN3A/B (BY SIMILARITY).
 FT DOMAIN 299 328 COILED COIL (POTENTIAL).
 FT DNA_BIND 437 482 SANT-A (POTENTIAL).
 FT DNA_BIND 624 669 SANT-B (POTENTIAL).
 FT DOMAIN 501 550 COILED COIL (POTENTIAL).
 FT DOMAIN 606 616 PRO-RICH.
 FT DOMAIN 2073 2077 CORNR BOX OF ID1.
 FT DOMAIN 2277 2281 CORNR BOX OF ID2.
 FT DOMAIN 58 64 POLY-GLN.
 FT DOMAIN 593 602 POLY-ALA.
 FT DOMAIN 1044 1047 POLY-PRO.
 FT DOMAIN 1713 1718 POLY-ALA.
 FT DOMAIN 1968 1979 POLY-SER.
 FT VARSPLIC 2333 2371 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 1952 1952 I -> T (IN REF. 2).
 FT CONFLICT 2090 2090 A -> P (IN REF. 2).
 SQ SEQUENCE 2453 AA; 270640 MW; 52208B40382F7E6A CRC64;

Query Match 3.7%; Score 127.5; DB 1; Length 2453;
 Best Local Similarity 18.6%; Pred. No. 6.3;
 Matches 123; Conservative 80; Mismatches 232; Indels 225; Gaps 28;

QY 3 SRKSKAALDTPOSESTFSTSLDSSAPSPARNLRSGRNILQPSSEKDRD-----HEK 56
 Db 103 SLESKRPRLE--QVSDSHFQRI--SAAVLPVHTLPEG---LRSSANAKKDPAGVKHEA 155
 QY 57 RSGEELAGRMGMGDANGCHLREGKEQEGVKMAIEGLARMERLQRA-----TKRQKKQ 110

QY 130 YHHKSTNAEEREAKPEVLKTHSKDVERAEI-----GVDDVVVKMEPAATN 174
 Db 1584 RTPRESTQKEKIPK--LLKVDSLNLQNSOLDNSVSDSDPIFSDPGFSCYSLEDS--- 1638
 QY 175 IIEPE-----DAQDAERGAARPAVNSSYLP-----LPWKGRGLGYACLNTYLRNAKPPIF 225
 Db 1639 -LSPEHNYNFDINTIGOTGFCFSYS-GSOFVPADQNLPOK-----FLSDAVQDLF 1686
 QY 226 SSRTCRMASIVDHRHPLQFEDEPEHIL-----KKNP 256
 Db 1687 PGOAIEKNEFLSHDN--QKDEKHHITTSASWIRSGTLSPEIFEKSTIDSNEENRRHNM 1744
 QY 257 DKSKEPQDELGHKFVQELGLANARDIVKMLCWNKEYGIRFLRLSSEMFPFASHPVHGYKL 316
 Db 1745 KNSFPHLTTTRSNSIMDSFCVQAAED-----CLSEKSLRNLSSSVSKEVFLSLPQPNNSDWI 1799
 QY 317 APFASEVLAERAGRAVAELGHLRTHPGQFTO--LGSPRKEVWESAIRDLEYHDELLSLLKL 375
 Db 1800 -----QGHTRKEMQOSLDSANTSETAILSSPDGELVDVACBDLE-----L 1839
 QY 376 PEQQRDAVMIIHMGQFGDKAATLEREKRNRYARLSOSCKNRLVLENDVGVTVHDLPLV 435
 Db 1840 YVSRNDML-----TPTDSSPRSTSSPSQS-KNGSFTPR-----TANILKPL 1881
 QY 436 C-----EELNIPWLDY-----HHNICDPALHREGTLDISDPKL--QERIAN----- 477
 Db 1882 MSPPSREEI-MATLLDHLSETIYQEPFCSPNPSDVPEKPREIGRLLMVETRLANDLAEF 1940
 QY 478 -----TWKR-----KGIKQKMHYSEPCDGAVTPRDRKRHRPR 509
 Db 1941 EGDFFSLEGLRLMKTAFSAMTONPRPGSPLRSGQGVVKNKSSNSP-----KMVEDKK 1991
 QY 510 VMTLP-PCPPDMDLM---IEAKDKEAVFELMRTFKLPGFEKINDMVPYDRDENRPAAP 565
 Db 1992 IVIMPCKCAPSRLQVWVLOAKEEYERSKKLPKT-KPTGVVKSANFNSSVNPDDKPVVP 2050
 QY 566 VK-----APKKKGGKRRKRTTDEEAAEFEEVDTAADVDK-- 600
 Db 2051 PKMDVSPCILPTAHTKEDVDNSQIALQAPTTCQSQTASESOMLPP--VASASDPEKDED 2108
 QY 601 -----APEGP-----KEVPEERAMGGPNRYVWP-----LCCEWLKPKKRE 638
 Db 2109 DDNYIYSYSSPDPVIPPWQOPISPDKALNGD-DRPSSPVEELPSLAFENFLKPIKDG 2167
 QY 639 VKKGKVPVEE 649
 Db 2168 IQKSPCSEPOE 2178

Search completed: January 15, 2002, 14:07:17
 Job time: 946 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:51:04 ; Search time 57.72 Seconds
(without alignments)
255.755 Million cell updates/sec

Title: US-09-724-296-36

Perfect score: 3482

Sequence: 1 MPSRKRAALDTPQSEST.....REVKKGVPEVEDEGEFDG 656

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_AA*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131.5	3.8	1098	4	US-08-923-992A-8
2	128	3.7	1805	1	US-07-853-913-2
3	121	3.5	1780	1	US-08-769-309A-5
4	121	3.5	1780	3	US-08-994-570-5
5	119.5	3.4	1128	4	US-08-923-992A-6
6	118.5	3.4	1164	4	US-08-923-992A-2
7	117.5	3.4	1104	4	US-08-923-992A-4
8	117.5	3.4	1588	5	PCT-US93-07261-11
9	117.5	3.4	1663	5	PCT-US93-07261-16
10	117.5	3.4	1872	1	US-08-188-582-14
11	117.5	3.4	1872	1	US-08-646-715-14
12	117.5	3.4	1893	1	US-08-188-582-11
13	117.5	3.4	1893	1	US-08-646-715-11
14	115	3.3	727	2	US-08-475-844-9
15	115	3.3	727	5	PCT-US95-08429-9
16	114.5	3.3	1618	1	US-07-853-913-4
17	113.5	3.3	1018	1	US-08-072-610-2
18	113.5	3.3	1018	2	US-08-719-822B-2
19	113.5	3.3	1018	4	US-09-092-458-2
20	113.5	3.3	1093	4	US-09-315-793-52
21	113	3.2	542	4	US-08-675-816-6
22	112	3.2	558	1	US-08-285-440-6
23	112	3.2	558	1	US-08-630-349-6
24	112	3.2	1068	4	US-09-085-199B-11
25	110.5	3.2	251	2	US-08-766-738-3
26	109.5	3.1	1213	4	US-09-413-814-79
27	109	3.1	410	2	US-09-018-628-14

28	109	3.1	410	3	US-09-273-378-14	Sequence 14, Appl
29	109	3.1	456	2	US-09-018-628-16	Sequence 16, Appl
30	109	3.1	456	3	US-09-273-378-16	Sequence 16, Appl
31	109	3.1	500	2	US-09-018-628-18	Sequence 18, Appl
32	109	3.1	500	3	US-09-273-378-18	Sequence 18, Appl
33	109	3.1	500	4	US-09-018-635-27	Sequence 27, Appl
34	108	3.1	1838	4	US-09-120-663-2	Sequence 2, Appl
35	107.5	3.1	251	2	US-08-766-738-1	Sequence 4, Appl
36	107	3.1	885	2	US-08-533-306A-4	Sequence 1, Appl
37	107	3.1	885	2	US-08-742-923A-4	Sequence 4, Appl
38	107	3.1	1788	2	US-08-962-284-2	Sequence 2, Appl
39	107	3.1	2237	1	US-08-455-543A-48	Sequence 48, Appl
40	107	3.1	2237	2	US-08-223-305C-48	Sequence 2, Appl
41	107	3.1	2337	3	US-08-713-118-2	Sequence 48, Appl
42	107	3.1	2337	4	US-09-452-007-2	Sequence 2, Appl
43	107	3.1	2339	1	US-08-455-543A-47	Sequence 47, Appl
44	107	3.1	2339	2	US-08-223-305C-47	Sequence 47, Appl
45	106	3.0	532	1	US-08-285-440-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1

US-08-923-992A-8

; Sequence 8, Application US/08923992A

; Patent No. 6280738

; GENERAL INFORMATION:

; APPLICANT: Tai, Joseph Y.

; APPLICANT: Blake, Milan S.

; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B

; TITLE OF INVENTION: Streptococcal Beta Antigens

; NUMBER OF SEQUENCES: 34

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

; STREET: 1100 New York Avenue, N.W., Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/923,992A

; FILING DATE: 05-SEP-1997

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/024,707

; FILING DATE: 06-SEP-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Esmond, Robert W.

; REGISTRATION NUMBER: 32,893

; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1098 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-923-992A-8

Query Match 3.8%; Score 131.5; DB 4; Length 1098;

Best Local Similarity 18.4%; Pred. No. 0.0042;

Matches 121; Conservative 94; Mismatches 259; Indels 183; Gaps 27;

QY 13 TPQSESTFTSLDSSAPARNLRSGRNIIQPSSEKDRDHEKRSGEELA--GRMMGKD 70

Db 31 TDGNNSSSELETTREIMPTDIDKAVEPVKTAGETSAHTGKREKQLOQWKNLKN 90
Qy 71 ANGHCLREGKQOEGVKMAI-----EGLARMERLQORATK-----RQK 108
Db 91 VDNTIL--SHEQKNEFKTKIDETNDSADALLENOFNETNKLHLIKOHEVEKDKKAKQ 148
Qy 109 KOLEDGIPVSVVSRPTAPYHHSKNAE--ERAKPEVLKTHSKDVERAEIGVDDVVK 167
Db 149 KTLKSDTKVD--LSNIDKELNHQSQEAGITNEDKDSMLK-KIEDIRKQAO-----197
Qy 168 MPAAATNIEPDQAADAAERPAVNSYLLPLPWKGRGLGYACLTLYLRNAKPPFPSS 227
Db 198 -----QPKKEDAEVK-----VREELGKLFSS 219
Qy 228 RTRMASIVDRHPLQFDEPEHLKKNPKDSKEPO--DE-----LGHKFVOELGLA 277
Db 220 TKAGL-----DOEQEHVKKETSEENTOKVDEHYANSLONIAOKSLEELDKA 267
Qy 278 NARDIVKMLCNWKEYGIRFLRLSSMFPFASHPVHGYKLAPFASEVLAEAGRVAAELGHR 337
Db 268 TTNEQATQV--KNQFLENAOKLKEIOPLIKET--NVKLYKAMSESLEQ---VEKELKH- 318
Qy 338 LTHFGQFTQGLSPKPEVSESAIRLDYHDELLSLKLPLPEQONRDVAVMIHMGQFGDKA 397
Db 319 --NSEANLEDLVAKSEIVREYEGKLNOSKNLPELKQLEEAHSAKSKOVVE---DFRKKF 373
Qy 398 ATLE-----RFRKRYA-----RLSQCKNRLVLENDVGMVTHDLLPVCCELNIP 442
Db 374 KTSQVTPKKRLKDLAANENNOQKIELTVSPENITVYEGEDVKFTV-----420
Qy 443 MVLDYHHNICFDPHAHLRGTLDISD-----PKQERANTWKBKGIKQKMHYSEPCDG 496
Db 421 -----TAKSDSKTTLDLDFSLTLTKYNPVSVDRISTNYKTNTDNHKT-----A 461
Qy 497 AVTPRRDRKRRPRVMTLPPCPDMDLMIKADKQAVFELMRTFKLPGEKINDMV---P 553
Db 462 EITIKNLKNSQTVL-----KAKDSDGNVVE--KTFTITVOKKEKQVPRTP 508
Qy 554 YRDDENRAPPVYKAPKKKGGKRRKTTDEEAPEEVDVTAADDVKDAPEGKYPE 610
Db 509 EQKDSKTEEKVP--QEPKSNKQLOELIKSAQOQLEKLEKATKELMEOPETPSN-PE 563

RESULT 2

US-07-853-913-2
Sequence 2, Application US/07853913
Patent No. 5338839
GENERAL INFORMATION:
APPLICANT: McKay, Ronald D.G.
APPLICANT: Lendahl, Urban
TITLE OF INVENTION: Nestin Expression As An Indicator of
TITLE OF INVENTION: Neuroepithelial Tumors
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/853,913
FILING DATE: 19920319
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/660,412
FILING DATE: 22-FEB-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,803
FILING DATE: 25-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/201,762
FILING DATE: 02-JUN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/180,548
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4641AAAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1805 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-853-913-2

Query Match 3.7%; Score 128; DB 1; Length 1805;

Best Local Similarity 18.8%; Pred. No. 0.018;

Matches 167; Conservative 102; Mismatches 287; Indels 334; Gaps 41;

Qy 13 TPOSESTSTLDSAPSPARNLRSGRNILOPSSEKORDHEKRSGE-----LAGRMGK 69
Db 685 SPEEDQACRPLQKNOEPLGYEAEQO-ILRLLEKESQESLRSPPEEDQAGRSLOK 743
Qy 70 -----DANGHCLKE--GKEQEGVKMAIEGLARMERLQORATKROKKOLEEDG---115
Db 744 ENOEPLGYEAEADOMLERLLEKESQESLSKSPEN-ORIGKPLERENOKSLRYLEENQETF 802
Qy 116 IPVSVVSVRPTAPYHHKSTNAERAKPEVLKTHSKDVERAEIGV-----DDVVK 167
Db 803 VPLSRNQR----PLRSLEVEEERQIRVPLEKVSQDSLSGLAEENQVPLRYLEEDDCIN 858
Qy 168 M-----EPAATNIIPEDAQDAARPAVNSYLLPLPWKGRGLGYAC 212
Db 859 KSLLKTHKSLGSLDRNGDSIIIPQSETOV--SLRPPPEEDQRI-----903
Qy 213 LNTYLRNAKPPFSSRTRMASIV-----DHRHPLQFEDPEPHHLKN 254
Db 904 VNHLEKESQE--FSRSSEEEQVNMERSLEGHENHESLSVVEKEQDMVESQLEKESQDSGKS 961
Qy 255 KPDKSKEP-----QDELGHKFFVOE-----273
Db 962 LEDESQETFGPLEKENAESLSLACQOQEBQKLEQETQOTLRVAGNEQMAVSPPEKVDPE 1021
Qy 274 -----LGLANARDIVKMLCNWKEYGIRFLR-LSSEMF-----304
Db 1022 LKPLGNQETARSIGKENQESVSL----KEGIVTKVKSLEIILEPLETAEDLERRK 1077
Qy 305 -----PFASHPVHGYKLAPFASEVLAEAGRVAAELGHRLTTHGQFTOLGSPKPEVSES 358
Db 1078 SIDTQEPLEWSTEVARETVPEPPGSLGSV-DENRETLTSLEKESQELSSLOKKNVET 1136
Qy 359 AIRDL-----EYHDELLSLIK--LPEQONR---DAY-----384
Db 1137 KVEDSQOQLQVEEGLOEQHQSLEKREVYKQELPSSGNQORWEDVVEKAVGQEAPLATTGV 1196
Qy 385 -----MIHMGQFGDKAATLREKRNRYARLSQCKNRLVLENDVVG--WTVHDLDPVCE 437
Db 1197 GTEDKAEHLRQGGEEBAAG-----ELLODIVGAWSLSGSEPP--K 1238
Qy 438 ELNIPM-VLDYHHNICFDPHAHLRGTLDI-----SDPKLOER-----IANTWK 480
Db 1239 EORVPAEALD-----NLEGALEVPVAOSMPTEVTERDEDRAGAGQDSIEVTIG 1287
Qy 481 RKGIKQKMHYSEPCDCAVTPRDRKRRPRVMTLPPCPDMDLM-----IEAKDKE-----530

Db 1288 LEAARTGLEQEVVGLDPR-----HFAREEAIPPSLGSESVKAKIAOGLGEGKPEKA 1343
QY 531 ----QAVFELMRT-----FKLPGFEKINDMVPYDR-----DDENRPPAPVKAP--KKK 572
Db 1344 GAIJDSGLFELPKTSSBALECOGHEESSEMEGWBEEBASLETSDEHGSADAPQPRPPEEED 1403
QY 573 KGGKRRT-----TD-----FEAAPEPVEDTAADDVKDAPE- 603
Db 1404 EGAALATAPGPKLLEPCSPILTDHAHELQQAEGIQAGHQPEAGSALERVEPEF 1463
QY 604 GPKEVEERAMGGPNRYNWPVLCGEWELKPKKREVKKGKVPVEVEDEGE 653
Db 1464 GLGEIPE-----GLQDW-----EEGRESEADDLGE 1489

RESULT 3
US-08-769-309A-5
; Sequence 5, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauck, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/769,309A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5741890and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1780 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-769-309A-5

Query Match 3.5%; Score 121; DB 1; Length 1780;
Best Local Similarity 18.3%; Pred. No. 0.079;
Matches 134; Conservative 95; Mismatches 283; Indels 220; Gaps 30;

QY 4 RKSAAALDTPQ-----SESTFTSTLDSSAPSPARNLRSGRNILQPSSEKDRDHEKR 57
Db 185 KDKTEKPTVOLLTVKDEGEAGACAGHDQPSLGAGEAASKSEPKSTKEPTEELTKR 244
QY 58 ---SGBELAGRMGMKDANGHCLREGKEQEGVKMATEGLARMERRLQRA TKRKQKLEEDG 115
Db 245 EQSHAEISPPAESGQAVECKEKEGKEK-----KEPSKSAESPT 284
QY 116 IPV-----SVVSRPPTAPYHHKSTNAEBREAKEPVLTHTSKDVEREAP-IGVDVVKMEP 170

Db 285 SPVTSETSTFKKFFTCQWAWRRKKTSTRPKPEDEVEASEKKKEQPEKVDTEDEGKAEV 344
QY 171 AATNIIPEDEQAADAAERAPPVAVNSY--LPLPWKGRGLGYACINTYLRNAKPPIFSSR 228
Db 345 ASEKLTASEQAHPQPAESAHEPRLSAEYKVELPSEBOVS-----CSOGPSEBKP 395
QY 229 TCRMASIIVDHRPL-QFDEDEPEHHLKNPKDSKE-----PODELGHKFFVQELG 275
Db 396 APLATEVEDEKLEVHOEEVAEVHVSTVEETEEQKTEVEETAGSVPAEELVGVMDABPOE 455
QY 276 LANARDIVKMLCWNKYGIRELRLSSEMFPFASHPVHGYKLAPFASEVLAEGAVAAELG 335
Db 456 AEPKELVKL-----KETCVSGEDPTQAGDLSPDEKVLKRPPEGVVSEV- 499
QY 336 HRLTHPGOFTQLGSPRKEVVESAIROLEYHDELLSLKLPEQONRDVMIHMGGOFGD 395
Db 500 -EMLSQERMKVQGSPLKLTST-----GLKKLSGKKQKG-----KRGG--GD 540
QY 396 KAATLEREPKRYARLSQCKNRLVLENDVGVTVHDLPLVCEELNTPWLDYHHHNCIFD 455
Db 541 E-----ESGE-----HTQVPA-----DSP-----D 555
QY 456 PAHLREGTLDISDPKLOERIANWKRGKIKOKMHYSEPCDA----- 497
Db 556 SOEQKKGESSASSPEPEEI--TCLXGLAEVQODGAEAGCATSDGKKRRCVTPWASF 613
QY 498 --VTPRRKRHRPVMTLPPCPDMLMIEAKRQAVFELMRTFKLPGEK-INDMV 554
Db 614 KMVTPK-KRVRRP-----SESDKEDEL-DKVKSATLSSTESTASENQEE 655
QY 555 DRDDENRPPVPKAPK-----KKGKKRRTTDEE-----AAPEEV 591
Db 656 MKGSVEEPK--EEPKRKYDTSVWEALICVSSKKARRRSSDEEGPKAMGGDHQA 713
QY 592 DTAADDVKDAPEG-----PREVPEERAMGGPNRYNWPVLCGEWELKPKK-----RE 638
Db 714 DEACKDKETGTDGILAGSOEHDPGQSSSPDOAGSPTEGE-----GVSTWESFKRLVTPRK 769
QY 639 VKKGVPVEED 650
Db 770 KSKSKLEKSED 781

RESULT 4
US-08-994-570-5
; Sequence 5, Application US/08994570
; Patent No. 6090929
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauck, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/994,570
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090929and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451

TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1780 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-994-570-5

Query Match 3.5%; Score 121; DB 3; Length 1780;
Best Local Similarity 18.3%; Pred. No. 0.079;
Matches 134; Conservative 95; Mismatches 283; Indels 220; Gaps 30;

QY 4 RKSAALDTPQ-----SSSTFSSLTSSAPSPARNLRSGRNILQPSSEKDRDHEKR 57
DB 185 KDKTERPDVTLTVKKDEGEAGAGDHQDPSLGAAGEAASKESEPKOSTEKPEETLKR 244
QY 58 --SGEELAGRMGKDANGHCLREGKEQEBEGVKMAIEGLARMERLQRAATRKQKQLEEDG 115
DB 245 BQSHAEISPPAESQAVECKEGERKE-----KEPSKSAESPT 284
QY 116 IPVP-----SVYSRPTAPYHHKSTNAEREAKEPVLKTHSKDVERAE-IGVDDVVKMEP 170
DB 285 SPVSETGTFKFTQGWAGRWKTKFRPKPEDEVEASEKKKEQEPKVDTEEDCKAEV 344
QY 171 AATNIEPEDAQAAGAAERPAVNSY--LPLPWGRIGYACLYLNKAPKPTFSSR 228
DB 345 ASEKLTAEOAHOPEAASAEHPRLSAEKVELPSEQVS-----GSQGPSSEKP 395
QY 229 TCRMASIVDHRHPL-QPEDEPEHLKKNPKDSKE-----PDDELGHKFKVQELG 275
DB 396 APLATEVEDEKIEVHQEVVAEVHVSVEERTEQKTEVEETAGSVPAEELVGMDAEPQE 455
QY 276 LANARDIVKMLCWNEKYGIRFLRLSSEMFPFASHPVHGYKLAPFASVLAEPAGVAAELG 335
DB 456 AEPKELVKL-----KETCVSGEDPTQCADLSPDEKVLSPKPPGVVSEV- 499
QY 336 HRLTHPCQFTQLGSPKREKVEVSAIROLEYHDELLSLKLPEQONRDVAMLIHMGQFGD 395
DB 500 -EMUSSQERMKVQSGPLKLTST-----GLKLSGKKKOG-----KRGG--GD 540
QY 396 KAATLERKRYARLSQCKNRLVLENDVGTWTVHDLPLVCEELNIPMLDYHHNICFD 455
DB 541 E-----ESGE-----HTQVPA-----DSP-----D 555
QY 456 PAHLREGTLDISPKLQRIANTWKRKGIKOKMHSYSEPCDGA----- 497
DB 556 SQEQKGESSASSPEPEI--TCLEKGLAEVQODGAEAGATSDGKKRKGVTTPWASF 613
QY 498 --VTPDRRKRPVMTLPPCPDMDLMEAKDEQAVFELMRFKLPGEK-INDMV 554
DB 614 KMTPEK-KRVRRP-----SESDREDEL-DKVKSAATLSSTESTASMOQE 655
QY 555 DRDDENRAPPYKAPK-----KGGKRRRTTDEE-----AAEPEV 591
DB 656 MKGSVEEPKP--EEPKRKVDTSVSWAELICVSSKKRARRSSDDEEGPKAMGGDHQA 713
QY 592 DTAADVDKADPG-----PREVPEERAMGPNRVVWPLGCEEWLKPCK-----RE 638
DB 714 DEAGDKETGTGDIILAGSQEHDPGQSSSPQAGSPTBGE-----GVSTWESFKRLVTPRK 769
QY 639 VKKGKVPPEVED 650
DB 770 KSKSKEEKSED 761

RESULT 5
US-08-923-992A-6
Sequence 6, Application US/08923992A

Patent No. 6280738
GENERAL INFORMATION:
APPLICANT: Tai, Joseph Y.
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.014.0001/RWE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1128 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-923-992A-6

Query Match 3.4%; Score 119.5; DB 4; Length 1128;
Best Local Similarity 18.8%; Pred. No. 0.054;
Matches 123; Conservative 95; Mismatches 266; Indels 169; Gaps 29;

QY 13 TPQSESTFSLTSSAPSPARNLRSGRNILQPSSEKDRDHEKRSGEELA---GRM--- 66
DB 26 TDGNNSSSSSELETTTMEIPTTDIKKA---VEPV-----EKTAGETSATDTCKREKQ 74
QY 67 -----MGKDANGHCLREGKKEQEBGVKMAI-----EGLARMERLQRAATK---RQKK 109
DB 75 LOOKNNLKNLNDVNTIL--SHEQKNEFKTIDETNDSADALLELENOFNETNLLHIKQHE 132
QY 110 QLEEDGIPVSVVSRPTAPYHHKSTNAEREAKEPVLKTHSKDVERAEIGVDDVVKM- 168
DB 133 EVEDK-----KKAKQKTLKQSDTKVDLSNIDKELNHQKSPVEKMA 173
QY 169 EP-AATN-----IIEPEDAQAAGAAERPAVNSYLPPLPWKGLGVACLYLNK 221
DB 174 EPKGTINEDKDSMLKTIEDIRKQAQADKKEDA-----EVKVRREL 214
QY 222 PPIFSSTCRMASIVDHRHPLQFQDEPEHLLKKNPKDSKEPQDELGHKFTVOELGLANARD 281
DB 215 GKLFSTKAGLDOEI--HEH-VKKETSSEENTQKVDHYANSIONLAOKSLEELDRAATTNE 272
QY 282 IVKMLCWNEKYGIRFLRLSSEMFPFASHPVHGYKLAPFASVLAEPAGVAAELGRLTH 341
DB 273 QATQV---KNOFLENAOKLKEMOPLIKET--NVKLYKAMSESELEO---VEKELKH---NS 321
QY 342 PGQFTQLGSPKREKVEVSAIROLEYHDELLSLKLPEQONRDVAMLIHMGQFGDKAATLE 401
DB 322 EANLEDLVAKSKELVREYEGKLNQSKNLPKLEBEAHSKLQVVVE---DFRKKFKTSE 378

Patent No. 6280738
GENERAL INFORMATION:
APPLICANT: Tai, Joseph Y.
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.014.0001/RWE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1128 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-923-992A-6

TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1780 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-994-570-5


```

QY 402 -----RFRNYA-----RLSQCKNRLVLENDVGVTHDLPVCEELNIPWVLD 446
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 379 QVTKRKRVRDLAANENNQOKIELTVSPENITVYEGEDVKFTV----- 421
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
QY 447 YHHNICFDPHAHLREGTLDIS-----PKLQRIANTWKRKGKIGKIMHYSEPCDGAATP 500
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 422 -----TAKSDSKTTLDFSDLLTKYNSVSDRISTNYKNTNDNHI-----AEITI 466
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
QY 501 RDRKRHRPRVWTLPPCPDMDLMEAKDEQAVFELMRTEKLPGEKINDMV---PYDRD 557
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 467 KNLKLNESQTVTL-----KAKDDSGNVVE---KTFTITVQKKEKQVPTPEQKD 513
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
QY 558 DENRPAVPKAPKGGKGRKRTTDEEAAPPEEVDTAADVDKDAPEGPREVPE 610
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 514 SKTEKVP-QEPKSNQKNOLEKSAQOELEKRAIKELMEQPEIPSN-PE 564
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|

RESULT 6
US-08-923-992A-2
: Sequence 2, Application US/08923992A
: Patent No. 6280738
: GENERAL INFORMATION:
: APPLICANT: Tai, Joseph Y.
: APPLICANT: Blake, Milan S.
: TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
: TITLE OF INVENTION: Streptococcal Beta Antigens
: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
: STREET: 1100 New York Avenue, N.W., Suite 600
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PC-DOS/MS-DOS
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/923,992A
: FILING DATE: 05-SEP-1997
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/024,707
: FILING DATE: 06-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Esmond, Robert W.
: REGISTRATION NUMBER: 32,893
: REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1164 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-923-992A-2

Query Match 3.4%; Score 118.5; DB 4; Length 1164;
Best Local Similarity 18.4%; Pred. No. 0.069;
Matches 122; Conservative 96; Mismatches 256; Indels 189; Gaps 29;

QY 13 TPQSSSTFSLDSSAPARNLRSGNITLQPSSEKDRDHEKRSGEELA---GRM--- 66
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 62 TDQGNSSSSELETTKMEIPTDIDKKA---VEPV-----EKTAGTATSATDTGKREKQ 110
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
QY 67 -----MGKDANGHCLREGKEQEGVKWAT-----EGLARMERRLORATK-----ROKK 109
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|

```

```

Db 111 LQWKNNLKNVDNFTL--SHEQKNEFKTKIDETNDSALLLENOFNETNRLLLHIQHE 168
QY 110 QLEEDGIPVPSVVSREPTAPYHUHSTNAEREAKPEVLKTHSKDVEREAEIGVDVVKM- 168
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 169 EVEKD-----KKAQOKTLKQSDTKVDLSNLDKELNHOKSOKVERMA 209
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
QY 169 -EPAATN-----IIPEDAQDAEAERPAVNSVYLPLPWKGRGLGYACILNTYLRAK 221
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 210 EQKGITNEDKSMLKKIEDIRKQAOQADKEDA-----EVKVRREL 250
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
QY 222 PPIESSRTRMASIIVDHRHPIQFEDPEPHLKNKPKDKSKEPO--DE-----IGHKPV 271
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 251 GKLFSTSTKAGL-----DQEIQEHVKKKETSSEENTQKVDEHYANSQNLAQKSL 298
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
QY 272 QELGLANARDIVKMLCWNEKYGIRFLRLSSEMPFASHPVHGKYLAPFASEVLAEAGRVA 331
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 299 BELDKATTNEQATV---KNQFLENAOKLKEIOLIKET--NVKLYKAMSESLEO---VE 350
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
QY 332 AELGHLRLTHPGQFTOLGSPKREVVESAIROLEYHDELLSLKLPEQONRDVMIHMG 391
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 351 KELKH---NSEANLEDLVAKSKEIVREYEGKLNQSKNLPKLEEAHSLKQVVE--- 404
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
QY 392 QFGDKAATLE-----REKRYA-----RLSQCKNRLVLENDVGVTHDLPV 436
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 405 DFRKKFKTSEQVTPKKRVKRDIAANENNQOKIELTVSPENITVYEGEDVKFTV----- 457
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
QY 437 EELNIPWLDYHHNICFDPHAHLREGTLDIS-----PKLQRIANTWKRKGKIMHY 490
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 458 -----TAKSDSKTTLDFSDLLTKYNSVSDRISTNYKNTNDNHI--- 497
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
QY 491 SEPCDGAATPDRRKRHRPRVWTLPPCPDMDLMEAKDEQAVFELMRTEKLPGEKIND 550
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 498 -----AEITIKNLKLNESQTVTL-----KAKDDSGNVVE---KTFTITVQKKEK 539
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
QY 551 MV---PYDRDDENRPAVPKAPKGGKGRKRTTDEEAAPPEEVDTAADVDKDAPEGPKE 607
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 540 QVPKTPKQKSKTEKVP-QEPKSNQKNOLEKSAQOELEKRAIKELMEQPEIPSN 598
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
QY 608 VPE 610
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 599 -PE 600
      :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|

RESULT 7
US-08-923-992A-4
: Sequence 4, Application US/08923992A
: Patent No. 6280738
: GENERAL INFORMATION:
: APPLICANT: Tai, Joseph Y.
: APPLICANT: Blake, Milan S.
: TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
: TITLE OF INVENTION: Streptococcal Beta Antigens
: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
: STREET: 1100 New York Avenue, N.W., Suite 600
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PC-DOS/MS-DOS
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/923,992A
: FILING DATE: 05-SEP-1997
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/024,707
: FILING DATE: 06-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Esmond, Robert W.
: REGISTRATION NUMBER: 32,893
: REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1164 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-923-992A-2

```

RESULT 8
CT-US93-07261-11
Sequence 11, Application PC/TUS9307261
GENERAL INFORMATION:
TITLE OF INVENTION: PfEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF

[illegible]

GENERAL INFORMATION:	PfEMP3 MALARIA ANTIGEN ANALOGS. ANTIBODIES AND USES THEREOF
TITLE OF INVENTION:	

Db 409 KERKMOELHKKKNYK---YQKLLEREKRE---NPDGEPLNTPREIHVIRP-----SDLMD 458
 QY 525 EAKDKEQA-VFELMRTFKLPGFEKINDWVYDRDDENRPPVYKAPKPKKGG-KRRRTTD 582
 Db 459 KGENKSAGHPFKYOPTKGLKEYEESHVSKDYQLEHE-----PPTKLPYEKGHVSREYQLD 514
 QY 583 EAAAEPEVDTAADDVKADPEGKPEVEERAMGGPNVYVWPLGCEEWLKPKKREVKKG 642
 Db 515 HEP-----PTKLPYEK---GHVSREY-----QLDNEVRDELPEYK 549
 QY 643 KVPPE--VEDEG-----EFD 655
 Db 550 HVSREYQLDNEGPTLKEYD 569

RESULT 9

PCT-US93-07261-16
 ; Sequence 16, Application PC/TUS9307261
 ; GENERAL INFORMATION:
 ; TITLE OF INVENTION: PFEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: John H. C. Blasdale
 ; STREET: One Giralda Farms
 ; CITY: Madison
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07940-1000
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: Apple Macintosh
 ; OPERATING SYSTEM: Macintosh 6.0.5
 ; SOFTWARE: Microsoft Word 5.1a
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/07261
 ; FILING DATE: 19930805
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/927,531
 ; FILING DATE: 07-AUG-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Blasdale, John H. C.
 ; REGISTRATION NUMBER: 31,895
 ; REFERENCE/DOCKET NUMBER: DX0288K
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201-822-7398
 ; TELEFAX: 201-822-7039
 ; INFORMATION FOR SEQ ID NO: 16:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1663 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; ORIGINAL SOURCE:
 ; ORGANISM: Plasmodium falciparum
 ; STRAIN: Malayan Camp
 ; PCT-US93-07261-16

Query Match 3.4%; Score 117.5; DB 5; Length 1663;
 Best Local Similarity 19.4%; Pred. NO. 0.15;
 Matches 132; Conservative 108; Mismatches 247; Indels 193; Gaps 33;

QY 41 RNLIOPSSSEKDRDHEKRSSEELAGRMGMKDANGHCLRECKDEEGVKNALIEGLARMERRL 100
 Db 18 RNALK---EKKLKEQKNDAQAKDITKES-----QDSSEKSLKEKVGNEALKEKEN 68
 QY 101 QRATKQKQKQLEDPGIPVSVVSREFTAPYHHKSTNABERAKEPV-----LKTHSKD 153
 Db 69 KETLK--KKELE-----NQKEKEKNKIKNDNEALKNKND 103
 QY 154 -----VEREALGVDDVVKMPEPATNIIPEDAQDAAGRAARPPAVNSSYLPPLPWGR 207
 Db 104 KODKKIVPKKPESEVKDLKEMELKEKEFIK-OHLKDYERKEKRRNWLRLSRDKLREI 162

QY 208 LGYACINLYLRNAKPIFFSSRTCRMASIVDRHRPL-----OFEDE-----PEHH 251
 Db 163 EOELKLNQALSAINELKERRASRPMVMKMGKMKDEVLEWIKKYDDEQAENKTKDEE 222
 QY 252 LKNKPDKKE-----PDDELG-----HKFVOELGLANAROTIVKMLCWNE 290
 Db 223 IKDKGDGYEIVETKEYGMRNALGELDEYBERYKRYLKDGECDLKDVEEKL---E 279
 QY 291 KYGIRFLRLSSSMFPFASHVHYKFLAPFASVLAFAEAGRAAELGHLRLTHPGQFTQLGS 350
 Db 280 ETGYCF---REKFP-----TTRILVKRRR-----NK 302
 QY 351 PRKEVVESAIRDL---EYHDELLSLKLPQOONRDAMVIMHMGQFGDKAATLERFKRN 406
 Db 303 EQKLLKEDKEKKLIAAEEPDE--KKIKLSDDDKVVVPVKNKSSFPDKFRAPDKKRTM 360
 QY 407 YARLSQSCNRLVLENDVGVTVHDLLPVCYELNIPWLDYHHHNCIFDPAHLREGTLDI 466
 Db 361 FYRUSELFP--IVPRKDE-----LAVCGD-SMDSKVNGKKLKSTFNPFRKRNKL-- 408
 QY 467 SDPKLQE--RIANTWKRKGIKOMHYSEPCGCAVTPDRRRKRRPRVMTLPCCPPDMDLMI 524
 Db 409 KERKMOELHKKKNYK---YQKLLEREKRE---NPDGEPLNTPREIHVIRP-----SDLMD 458
 QY 525 EAKDKEQA-VFELMRTFKLPGFEKINDWVYDRDDENRPPVYKAPKPKKGG-KRRRTTD 582
 Db 459 KGENKSAGHPFKYOPTKGLKEYEESHVSKDYQLEHE-----PPTKLPYEKGHVSREYQLD 514
 QY 583 EAAAEPEVDTAADDVKADPEGKPEVEERAMGGPNVYVWPLGCEEWLKPKKREVKKG 642
 Db 515 HEP-----PTKLPYEK---GHVSREY-----QLDNEVRDELPEYK 549
 QY 643 KVPPE--VEDEG-----EFD 655
 Db 550 HVSREYQLDNEGPTLKEYD 569

RESULT 10

US-08-188-582-14
 ; Sequence 14, Application US/08188582
 ; Patent No. 5534410
 ; GENERAL INFORMATION:
 ; APPLICANT: Tjian, Robert
 ; APPLICANT: Comai, Lucio
 ; APPLICANT: Dynlacht, Brian D.
 ; APPLICANT: Hoey, Timothy
 ; APPLICANT: Ruppert, Siegfried
 ; APPLICANT: Tanese, Naoko
 ; APPLICANT: Wang, Edith
 ; APPLICANT: Weinzierl, Robert O.J.
 ; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
 ; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
 ; NUMBER OF SEQUENCES: 36
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FLEHR, ROHBACH, TEST, ALBRITTON & HERBERT
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/188,582
 ; FILING DATE: 28-JAN-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Osman, Richard A.
 ; REGISTRATION NUMBER: 36,627

Db 1728 POASVLYEDLLMSEGEDDEEDAGSDGDNPFSAIQLSESGSDVGGG----- 1777

Qy 628 CEWLKPKKREVKKGVPEVEDE 651

Db 1778 -----IRPKQPRMLQENTRMDMENE 1797

RESULT 13

US-08-646-715-11
; Sequence 11, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Neoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; NUCLEIC ACIDS ENCODING TAFS AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646.715
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1893 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-646-715-11

Query Match 3.4%; Score 117.5; DB 1; Length 1893;
Best Local Similarity 18.5%; Pred No. 0.18;
Matches 149; Conservative 107; Mismatches 275; Indels 273; Gaps 38;
Qy 16 SESSTFSLDSSAPSPANLRSGRNI-----LQPSKDRDHEKRSGEEL-----A 63
Db 1099 SSTVLSTDTDSSAEDS-DFEEMGKNINLQNKTSQSLSREREQERKELQRMLLAA 1157
Qy 64 G-----RMGKDANGCHLR-----EGKE----- 81
Db 1158 GSAAGNNHDDTASVTSLSNSATGRCILKIYRTFRDECKEYVRCETVRKPAIDAYVR 1217
Qy 82 -----QEEGV-KMAI-----EGLARMERRLQRTKROKKQLEEDGIPVP----- 119

Db 1218 IRTTDEEFTKAFALFDEQHREEMKERRRLOEQRLRLKRNQEKLKGPPEKPKMKKE 1277
Qy 120 -----SVVSREPT---APYHHKSTNAEEREAKPEVLKTHSKDVEREAEI---GV 162
Db 1278 RPDKLKCGACCAIGHMRTNKFPLYQ-TNA---PPSNPVAMTEOELEKTVIHNDN 1333
Qy 163 DDVVKMEPAATNIIPEDAQDAERGAARPPAVN--SSVLPILPWKGRGLGYACLNITYLRNA 220
Db 1334 EELIKVE--GTKIVLGKQLIESADEVRRKSLVKPKPKQOLPPKKARRVGTTVHCDYLRNP 1391
Qy 221 KPLIFSSRTCRMASIVDHRHPL--QFEDRPEHLKKNKDKSKEPQDELGHKFPV-OELGIA 277
Db 1392 HKSIIHRRITDPWVTLSSIIESTIINDMRDLPNITYPHPTPNAAKWKDY--YKLIIRPMDLQ 1449
Qy 278 NARDIVKMLCWNEKYGIRFLRLSSEMFPFASHPVHGKYLAPFASVLAERAGRAAEELGHR 337
Db 1450 TLKENVRK-----RIYPSKEEFREH-----LELIYK-----N 1476
Qy 338 LTHPGQFTQLGSPKKEVVESAIRDLEYHDELLSLKLPEOONRDVAMIHHMGOGFGDKA 397
Db 1477 SATYNG-----PKHSLTQISQMLDLDCDE-----KLKEXE-----DKL 1509
Qy 398 ATLERFKRYARLSOSCKNKLWLENDVGVGT-----VHDLPLVCEELNIPMV 444
Db 1510 ARLEK-----AINPLLDDDDQVAFSFLDNIVTQKMMAYFDSWPPEHPVYNKFPV 1558
Qy 445 LDYHHHNICFPAHLREGTLDISDPKLOER-----IANTWKRKG-----IKQRMH 489
Db 1559 PDY--YKVIIVNPMDELTIRKNSKHKYQSKSEFLDDVNLILANSVKYNGPESQYTKTAQE 1616
Qy 490 YSEPCDGAVTPRDRKRRPRVMTLPCCPPMDMLMIEAKDKEQAVFELM-----RTFKL 542
Db 1617 IVNVGYQTLTEYDEH-----LTQLEKDICTAKKAA-LEEALESLEDPMTPTGPTPOP 1667
Qy 543 PGEKINDWVPYDRD-----DENRAP--PVKAPKKK-----KGKRRKRTTDECA--- 585
Db 1668 PDLYDTNTSLNSRDSASFQDSNMSVLDIPSATPEKQVTQEGDGDGLADEEGTVQO 1727
Qy 586 -----AEPPEVDTAADVDKDAPEGPKVEPE-----EERAMGGPYNRVYWPGLG 627
Db 1728 PQASVLYEDLLMSEGEDDEEDAGSDGDNPFSAIQLSESGSDVGGG----- 1777
Qy 628 CEWLKPKKREVKKGVPEVEDE 651
Db 1778 -----IRPKQPRMLQENTRMDMENE 1797

RESULT 14

US-08-475-844-9
; Sequence 9, Application US/08475844
; Patent No. 5972643
; GENERAL INFORMATION:
; APPLICANT: Lobanenko, Victor V.
; APPLICANT: Neiman, Paul E.
; APPLICANT: Klenova, Elena M.
; APPLICANT: Goodwin, Graham H.
; APPLICANT: Filippova, Galina N.
; APPLICANT: Collins, Steven J.
; TITLE OF INVENTION: CTCF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourile and Crew
; STREET: One Market Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: PC compatible
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:50:00 ; Search time 120.01 Seconds
(without alignments)
404.900 Million cell updates/sec

Title: US-09-724-296-36

Perfect score: 3482

Sequence: 1 MPSRKSAAALDTPQSEST.....REVKKGVPEVEDEGEFDG 656

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3472	99.7	656	17 AAW01618	Neurospora crassa
2	3292	94.5	626	21 AAY44502	N. crassa delta228
3	3292	94.5	626	21 AAY44503	B. subtilis delta2
4	1011.5	29.0	828	21 AAY44498	GST signal peptide
5	963	27.7	600	21 AAY44500	GST signal peptide
6	960	27.6	371	21 AAY44499	S. pombe delta228-
7	262.5	7.5	294	21 AAY44505	D. radiodurans del
8	160.5	4.6	1341	21 AAY85657	Human Acinus L pro
9	143.5	4.1	914	22 AAB94420	Human protein sequ
10	139	4.0	747	21 AAB19002	A human microtubul
11	139	4.0	747	21 AAB29444	Human foetal liver

12	137	3.9	3256	21	AAV50976	Human cell cycle p
13	136	3.9	2665	22	AAV14533	Peptide #967 encod
14	136	3.9	2665	22	AAV26950	Peptide #987 encod
15	136	3.9	2665	22	AAV02259	Peptide #941 encod
16	135	3.9	1590	22	AAV73488	Mouse Rim2, a nove
17	135	3.9	3266	21	AAV42491	Human ORFX ORF2255
18	134.5	3.9	1117	22	AAV39632	Human polypeptide
19	134	3.8	2663	22	AAV39097	Human polypeptide
20	134	3.8	2688	22	AAV40883	Human polypeptide
21	133.5	3.8	583	21	AAV85659	Human Acinus S' pr
22	133.5	3.8	980	21	AAV18294	Plasmodium falci
23	131.5	3.8	1129	21	AAV84462	Amino acid sequenc
24	129.5	3.7	1596	18	AAV31347	Rat tumour suppres
25	128	3.7	617	22	AAV16458	Peptide #2892 enco
26	128	3.7	617	22	AAV04187	Peptide #2869 enco
27	128	3.7	1805	13	AAV27204	Rat nestin. Rattu
28	128	3.7	1805	15	AAV60126	Rat nestin. protein
29	128	3.7	2819	22	AAV35408	Human 07CG27 gene
30	127.5	3.7	2453	21	AAV12454	HNCR protein sequ
31	126.5	3.6	1272	22	AAV84881	Murine protein, SE
32	126.5	3.6	1272	22	AAV84883	Murine protein, SE
33	126.5	3.6	1490	22	AAV65644	Novel protein kina
34	126.5	3.6	2243	22	AAV84884	Murine protein, SE
35	125.5	3.6	1072	22	AAV70871	C. albicans apoptos
36	125.5	3.6	2441	21	AAV18161	Plasmodium falci
37	125	3.6	1382	18	AAV31867	Human metastasis-a
38	123.5	3.5	567	20	AAV88788	Polypeptide fragme
39	122.5	3.5	2518	21	AAV40574	Human ORFX ORF338
40	122	3.5	530	21	AAV26284	Arabidopsis thalia
41	122	3.5	530	21	AAV46611	Arabidopsis thalia
42	122	3.5	567	21	AAV46610	Arabidopsis thalia
43	121.5	3.5	568	21	AAV85658	Human Acinus S' pro
44	121.5	3.5	2096	21	AAV41592	Human ORFX ORF1356
45	121	3.5	1780	19	AAV53863	Human gravlin polyp

ALIGNMENTS

RESULT 1

AAW01618	ID	AAW01618 standard; Protein: 656 AA
XX	AC	AAW01618
XX	DT	20-AUG-1997 (first entry)
XX	DE	Neurospora crassa DNA repair enzyme.
XX	DE	DNA repair enzyme; cyclobutane-type dimer; (6-4)-bound product;
KW	KW	exposure; UV light; ultraviolet; suntan cream; prevention; treatment;
KW	KW	skin cancer.
XX	OS	Neurospora crassa.
XX	PN	JP08266275-A.
XX	PD	15-OCT-1996.
XX	PF	29-MAR-1995; 95JP-0094137.
XX	PR	29-MAR-1995; 95JP-0094137.
XX	PA	(SHTS) SHISEIDO CO LTD.
XX	DR	WPI; 1996-512663/51.
XX	DR	N-PSDB; AAV58286.
XX	PT	DNA repair enzyme and related DNA - specifically recognises
XX	PT	cyclobutane-type dimer and (6-4)-bound product induced by exposure
XX	XX	to UV light
XX	XX	Disclosure; Page 10-13; 18pp; Japanese.

```
XX The present sequence shows a DNA repair enzyme which specifically
CC recognises a cyclobutane-type dimer and (6-4)-bound product induced,
CC respectively from TT and TC sequences in DNA, by exposure to UV light.
CC The enzyme cleaves these sequences at the 5'-side. The DNA repair enzyme
CC can be used in, e.g. suntan cream for prevention and treatment of skin
CC cancer. Conventional endonuclease acts on either the cyclobutane-type
CC dimer or the (6-4)-bound product in damaged DNA, while the enzyme
CC of the invention can act on both of them.
XX
SQ Sequence 656 AA:

Query Match 99.7%; Score 3472; DB 17; Length 656;
Best Local Similarity 99.7%; Pred. No. 1.5e-295;
Matches 654; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MFSRKSAAALDTPQSESTFSSTLSSAPSPARNLRSGRNILQPSSEKDRDHEKRSGE 60
DB 1 mpsrksaaaltpqsesstfsstlssapsparnlrsgrnilqpssekdrdhekrsg 60

QY 61 ELAGRMGKDANGCHLREGKEGEGVKMAIEGLARMERRLQPAKTRQKKOLEEDGIPVPS 120
DB 61 elagrmgkdangchclregkegegvmkieglarmerrlqpaktrqkkoleedgipvps 120

QY 121 VYSRPTAPYHHKSTNAEREAKEPVLTHTSKDVEREAIEIGVDDVVKMEPAATNIIEPED 180
DB 121 vvsrptapyhhkstnaereakepvlthtskdvereaeigvddvvkmeapaatniieped 180

QY 181 AQDAERGAARPPAVNSSYLPPLPWKGRIGYACNLNYLRNAKPPFSSTTCRMASIVDHRH 240
DB 181 aqdaergaarppavnsyplpwwkgrigyacnlnylrnakppfssstcrmasivdhrh 240

QY 241 PLQFEDEPEHLLKNPKDKSEPDQLGHKFVOELGLANARDIVKMLCWNKYGIRFLRLS 300
DB 241 plqfedepehllknpkdksepdqlghkfvoelglanardivkmlcwnkygirflrls 300

QY 301 SEMFPFASHPVHGYKLAFFASEVLAEGRAVAELGHLRTHHPGQFTQLGSPRKEVVEAI 360
DB 301 semfpfashpvhgyklaffasevlaeagrvaaelghlrlthhpqgftqlgsprkevvesai 360

QY 361 RDLEYHDELLSLKLPEQONRDVMI IHMGQFGDKAATLERFKRNYARLSOSCKNRLVL 420
DB 361 rdleyhdehllslklpeqonrdavmiihmgoFGDKAATlerfkrnyarlsoscknrlvl 420

QY 421 ENDVGWTVHDLPLVCEELNIPMLVDYHHNITCFDPAHLREGTLTIDSPKQERIANTWK 480
DB 421 endvgwvthdlplvceelnipmvl dyhhnitcfdpahlregtltidspkqeriantwk 480

QY 481 RKGIKOMHYSEPCDGAFTPRDRKRHRPRVMTLPPCPDMDLMTAKDKEQAVFELMRTF 540
DB 481 rkgikomhysepcdgaftprdrkrhrprvmtlppcpdmdlmtakdkeqavfelmrft 540

QY 541 KLPQFEKINDMVPYDRDNRNRPAPPVKA PKKKGGKRRKTDEEAPEEVDVTAADDVKD 600
DB 541 klpqfekindmvpdydrdnrnrapppvka pkkkggkrrktdeeaapeevdvtaaddvkd 600

QY 601 APEGKVPVEERAMGGPNRVYWPGLGCEEWLKPKKREYKKGKVPVEEVEGEFDG 656
DB 601 apegkvpveeramggpnrvywplgceewlkpkkreykkgkvpveevegefdg 656
```

```
RESULT 2
AAAY44502
ID AAAY44502 standard; Protein: 626 AA.
XX
AC AAAY44502;
XX
DT 27-MAR-2000 (first entry)
XX
DE N. crassa delta228-UV damage endonuclease.
XX
KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;

glutathione-S-transferase signal peptide; uvel+ gene product;
UV irradiation; DNA damage; UV radiation damage; photoproduct;
basic site; platinum diaduct; mismatched nucleotide pairing;
nucleotide alkylation; skin cancer.
Neurospora crassa.
WO9963828-A1.
XX
PD 16-DEC-1999.
XX
XX 08-JUN-1999; 99WO-US12910.
XX
PR 08-JUN-1998; 98US-0088521.
PR 18-MAY-1999; 99US-0134752.
XX
PA (UYEM-) UNIV EMORY.
XX
XX Boetsch PW, Kaur B, Avery AM;
XX
DR WPI; 2000-116417/10.
XX
XX A new truncated ultraviolet damage endonuclease for treatment of skin
PT cancers -
XX
XX Claim 16; Page 59; 133pp; English.
XX
PS The present sequence is delta228-UV damage endonuclease from N. crassa.
CC Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
CC uvel+ gene product. This is expressed in frame with a GST leader sequence
CC to generate a fusion protein. This provides stable endonuclease fragments
CC for cleaving a double-stranded DNA molecule that has a distorted
CC structure resulting from UV radiation damage, a photoproduct, an abasic
CC site, mismatched nucleotide pairing, platinum diaduct, an intercalated
CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
CC for internal or topical application and as a therapeutic agent for skin
CC cancers.
XX
SQ Sequence 626 AA;

Query Match 94.5%; Score 3292; DB 21; Length 626;
Best Local Similarity 95.4%; Pred. No. 8.5e-280;
Matches 626; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 MFSRKSAAALDTPQSESTFSSTLSSAPSPARNLRSGRNILQPSSEKDRDHEKRSGE 60
DB 1 mpsrksaaaltpqsesstfsstlssapsparnlrsgrnilqpssekdrdhekrsg 60

QY 61 ELAGRMGKDANGCHLREGKEGEGVKMAIEGLARMERRLQPAKTRQKKOLEEDGIPVPS 120
DB 61 elagrmgkdangchclregkegegvmkieglarmerrlqpaktrqkkoleedgipvps 120

QY 121 VYSRPTAPYHHKSTNAEREAKEPVLTHTSKDVEREAIEIGVDDVVKMEPAATNIIEPED 180
DB 121 vvsrptapyhhkstnaereakepvlthtskdvereaeigvddvvkmeapaatniieped 180

QY 181 AQDAERGAARPPAVNSSYLPPLPWKGRIGYACNLNYLRNAKPPFSSTTCRMASIVDHRH 240
DB 181 aqdaergaarppavnsyplpwwkgrigyacnlnylrnakppfssstcrmasivdhrh 240

QY 241 PLQFEDEPEHLLKNPKDKSEPDQLGHKFVOELGLANARDIVKMLCWNKYGIRFLRLS 300
DB 241 plqfedepehllknpkdksepdqlghkfvoelglanardivkmlcwnkygirflrls 300

QY 301 SEMFPFASHPVHGYKLAFFASEVLAEGRAVAELGHLRTHHPGQFTQLGSPRKEVVEAI 360
DB 289 ---fpfashpvhgyklaffasevlaeagrvaaelghlrlthhpqgftqlgsprkevvesai 345

QY 361 RDLEYHDELLSLKLPEQONRDVMI IHMGQFGDKAATLERFKRNYARLSOSCKNRLVL 420
DB 346 rdleyhdehllslklpeqonrdavmiihmgoFGDKAATlerfkrnyarlsoscknrlvl 405
```

QY 421 ENDDVGWTVHDLVPVCEELNIPMWLDYHHHNCIFDPAHLREGTLDISDPKLOBRIANTWK 480
 Db 406 enddvgwtvhdllvpvceelnipmvdyyhhnncifdpahiregtldisdpklqiriantwk 465
 QY 481 RKGTKQKMHYSEPCDGAVTDRDRKHPRVMTLPCCPPMDLMEAKDKEQAVFELMRTF 540
 Db 466 rkgtkqkmysepcdgavtprdrkrhprvmtlpccppmdlmeakdkeqavfelmrtf 525
 QY 541 KLPGFEKINDMVPYDRDENRPPVKKPKKKGGKRRKTTDEEAPEPEEVDTAADDVKD 600
 Db 526 klpgfekindmvpdydrdenrppvpkpkkkggkrrkttddeapepeevdtaddvkd 576
 QY 601 APEGKPEVPEERAMGGPNRVYWPGLGCEEWLKPKKREVKKGKVPPEVEDEGEFDG 656
 Db 577 -----evpeeeramgpnrvywpplgceewlkpkkrevkkgkvppevedegefdg 626

RESULT 3

AA444503
 ID AAY44503 standard; Protein: 626 AA.
 AC AAY44503;
 XX
 XX 27-MAR-2000 (first entry)
 DT
 XX
 XX B. subtilis delta228-UV damage endonuclease.
 DE
 DE Delta228-UVDE: ultraviolet damage endonuclease; GST signal peptide;
 KW glutathione-S-transferase signal peptide; uvel+ gene product;
 KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
 KW abasic site; platinum diaduct; mismatched nucleotide pairing;
 KW nucleotide alkylation; skin cancer.
 XX
 OS Bacillus subtilis.
 XX
 XX W09963828-AL.
 PD
 XX 16-DEC-1999.
 XX
 XX 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0089521.
 PR 18-MAY-1999; 99US-0134752.
 XX
 PA (UYEM-) UNIV EMORY.
 XX
 XX Doetsch PW, Kaur B, Avery AM;
 XX
 XX WPI; 2000-116417/10.
 DR
 XX
 PT A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers -
 PT
 PS Claim 16; Page 59; 133pp; English.
 XX

CC The present sequence is delta228-UV damage endonuclease from B. subtilis.
 CC Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
 CC uvel+ gene product. This is expressed in frame with a GST leader sequence
 CC to generate a fusion protein. This provides stable endonuclease fragments
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, platinum diaduct, an intercalated
 CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 SQ Sequence 626 AA;

Query Match 94.5%; Score 3292; DB 21; Length 626;
 Best Local Similarity 95.4%; Pred. No. 8.5e-280;
 Matches 626; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 MPSRKSAAALDTPOSESTYSSSTLDSSAPARNLRSGRNLOPSSSEKORDHEKRSGE 60
 Db 1 mpsrksaaaltdtpsesstysstldssapsarnlrrsgnllqpsssekdrrdhekrsg 60
 QY 61 ELAGRMGKDKANGHCLREGKEQEGVKMAIEGLARMERRLORATRKOKKOLEEDGIPVPS 120
 Db 61 elagrmmgkdanghclregkeqeeyvkmaieglarmerrlgratrkqkqleedgipvps 120
 QY 121 VVSRPTAPYHHKSTNAEEREAKEPVLKTHSKDVEREAEIGVDDVYVKMEPAATNIIEPD 180
 Db 121 vvsrptapyhhkstnaeereakepvltkthskdvereaeigvddvvyvkmepaatniiepd 180
 QY 181 AQDAEAERGAARPAVNSSYLPLPWKGRGLGYACLNITYLNNAKPPPISSRTCRMASIVDRH 240
 Db 181 aqdaeragaarppavnsyylplpwkgrlgyacintyltnakppifssrtcrmasivdrh 240
 QY 241 PLQFEDEPEHHLKPKDKSKPEQDELGHKFVQELCLANARDIVKMLCWNEKYGIRFLRLS 300
 Db 241 plqfedepehlkpkdkskpeqdelghkfvqelglanardivkmlcw----- 288
 QY 301 SEMPPFASHPVHYGKLAPFASEVLAEGRAVAELGHRLLTTHPGQFTQLGSPRKEVESAI 360
 Db 289 ---fpashpvhgylapfasevlaeagrvaaelghrltthpgqftqlgspkrkewesai 345
 QY 361 RDLEYHDELLSLKLPQONRDVMIIMHGGQFGDKAATLERFRKRYARLSQSKNRLVL 420
 Db 346 rdleyhdelisllkpeqonrdavmlihmggqfgdkaatlerfrkryarlsqsknrlvl 405
 QY 421 ENDDVGWTVHDLVPVCEELNIPMWLDYHHHNCIFDPAHLREGTLDISDPKLOBRIANTWK 480
 Db 406 enddvgwtvhdllvpvceelnipmvdyyhhnncifdpahiregtldisdpklqiriantwk 465
 QY 481 RKGTKQKMHYSEPCDGAVTDRDRKHPRVMTLPCCPPMDLMEAKDKEQAVFELMRTF 540
 Db 466 rkgtkqkmysepcdgavtprdrkrhprvmtlpccppmdlmeakdkeqavfelmrtf 525
 QY 541 KLPGFEKINDMVPYDRDENRPPVKKPKKKGGKRRKTTDEEAPEPEEVDTAADDVKD 600
 Db 526 klpgfekindmvpdydrdenrppvpkpkkkggkrrkttddeapepeevdtaddvkd 600
 QY 601 APEGKPEVPEERAMGGPNRVYWPGLGCEEWLKPKKREVKKGKVPPEVEDEGEFDG 656
 Db 577 -----evpeeeramgpnrvywpplgceewlkpkkrevkkgkvppevedegefdg 626

RESULT 4

AA444498
 ID AAY44498 standard; Protein: 828 AA.
 XX
 AC AAY44498;
 XX
 XX 27-MAR-2000 (first entry)
 DT
 XX
 XX GST signal peptide and S. pombe UVDE fusion protein.
 DE
 DE GST signal peptide; ultraviolet damage endonuclease; UVDE; Uvelp;
 KW UV irradiation; DNA damage; UV radiation damage; fusion protein;
 KW skin cancer; glutathione-S-transferase.
 XX
 OS Schizosaccharomyces pombe.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FT Peptide 1..229
 FT Protein /label= GST_signal_peptide
 FT 230..828
 FT /note= "S. pombe UVDE"
 FT Misc-difference 11
 FT /note= "Encoded by aaaa"
 FT Misc-difference 85
 FT /note= "Encoded by gtt"
 FT Misc-difference 147
 FT /note= "Encoded by gtt"


```

XX      Query Match      27.7%; Score 963; DB 21; Length 600;
SQ      Best Local Similarity 41.4%; Pred. No. 1.2e-75;
        Matches 218; Conservative 60; Mismatches 125; Indels 124; Gaps 13;

QY 145 PVLKTHSVDREAEIGVDDVVKMEPAATNIEPEDAQDAEAGARPAPVANSYL----- 200
DB 176 pklvckkrie-----aipqdkylsksyiawplgqatfgggdhhp--ksdhivprg 228

QY 201 -----PLPWKGRIGYACLNLYLRNAKPPPISSRTCRMASTVDHRHPLQFE 245
DB 229 sddhapremfdcdkpiwgrlgyacnltlrmkervfcsrtcrtt----- 278

QY 246 DEPEHLKKNPKDSKEPDDELGHKFVOEGLANARDIVKMLCWNEYGIRFLRLSSEMPF 305
DB 279 -----qrdglesvkqigtqnvldliklvewnhnfhfmrsvssdlfp 320

QY 306 FASHVHYKLPAPFASVLAEGRAAEGLHRLTTHPGOFTQLGSPRKEVVEAIRDLEY 365
DB 321 fashakygyl-efagshleevgklankynhriltmhpqyqtqiasprevvvdsairdlay 379

QY 366 HDELLSLKLPQOONRDVAMVIHMGQFGDKAATLERFKRNYARLSQSCKNRLVLENDV 425
DB 380 hdeillsrmklineqldavliihlgtfegkktldfrknygrlsdsvkarlvlemdv 439

QY 426 GWTVDHLLPVCBEELNPWLDYHHNICFDPAPHLREGTLDISPKLOERIANTWKRGK 485
DB 440 swsvqdlpicqelnplvldwhhnlv--pgtlregsidlm--pliptiretwtkrkit 495

QY 486 QKMHYSEPCD-GAVTPDRRKHPRVMTLPPCPPDMDLMEAKDQVAFELMRTFKLP 544
DB 496 qkqhysesadptaisgmkrhdsrvidfpdptndlmieakeqavfelcrryel-- 553

QY 545 FEKINDWVPYDRDDENRAPPVKAPKKKGKRRKRTTDEEAPEEVDTAADDVKAPEG 604
DB 554 -----qnpccp-----leimgpe-----yqdtldg--- 573

QY 605 PKEVPEERAMGPGYNRVVWPLGCEEWLKPKEVKKGKVPVEVEDE 651
DB 574 -----yypgaekrltarkrrsrk-----eeveed 598

RESULT 6
AAY44499
ID AAY44499 standard; Protein; 371 AA.
AC AAY44499;
XX
XX
XX 27-MAR-2000 (first entry)
XX
XX S. pombe delta228-UV damage endonuclease.
XX
XX Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
KW glutathione-S-transferase signal peptide; uvel+ gene product;
KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
KW abasic site; aplatium diaduct; mismatched nucleotide pairing;
KW nucleotide alkylation; skin cancer.
XX
OS Schizosaccharomyces pombe.
XX
XX WO9603828-A1.
XX
XX 16-DEC-1999.
XX
XX 08-JUN-1999; 99WO-US12910.
XX
XX 08-JUN-1998; 98US-0088521.
XX 18-MAY-1999; 99US-0134752.
XX
XX (UYEM-) UNIV EMORY.
PA

```

Doetsch PW, Kaur B, Avery AM;
WPI: 2000-116417/10.
N-PSDB; AAZ29859.
A new truncated ultraviolet damage endonuclease for treatment of skin cancers -
Claim 13; Page 53; 133pp; English.
The present sequence is delta228-UV damage endonuclease. Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the S. pombe uvel+ gene product. This is expressed in frame with a GST leader sequence to generate a fusion protein. This provides stable endonuclease fragments for cleaving a double-stranded DNA molecule that has a distorted structure resulting from UV radiation damage, a photoproduct, an abasic site, mismatched nucleotide pairing, aplatium diaduct, an intercalated molecule or alkylation of a nucleotide. Uvelp can be used in compositions for internal or topical application and as a therapeutic agent for skin cancers.

Sequence 371 AA;
Query Match 27.6%; Score 960; DB 21; Length 371;
Best Local Similarity 45.4%; Pred. No. 1.1e-75;
Matches 205; Conservative 54; Mismatches 95; Indels 98; Gaps 10;

QY 201 PLPWKGRIGYACLNLYLRNAKPPPISSRTCRMASTVDHRHPLQFEDEPEHLKKNPKDSK 260
DB 15 pipwgrlgyacnltlrmkervfcsrtcrtt----- 49
QY 261 EPQDELGHKFVOEGLANARDIVKMLCWNEYGIRFLRLSSEMPFASHVHYKLPAPFA 320
DB 50 ---qrdglesvkqigtqnvldliklvewnhnfhfmrsvssdlfpashakygyl-efa 105
QY 321 SEVLAEACRAAEGLHRLTTHPGOFTQLGSPRKEVVEA:RDLEYHDELLSLKLPQOON 380
DB 106 qshleevgklankynhriltmhpqyqtqiasprevvvdsairdlayhdeillsrmklineq 165
QY 381 RPAVMTIHMGGQFGDKAATLERFKRNYARLSQSCKNRLVLENDVGVTVHDLPLVCEELN 440
DB 166 kdavliihlgtfegkktldfrknygrlsdsvkarivlendvsvsqdlpicqeln 225
QY 441 IPWLDYHHNICFDPAPHLREGTLDISPKLOERIANTWKRGKIQKMHYSEPCD-GAVT 499
DB 226 ipvldwhhnlv--pgtlregsidlm--pliptiretwtkrkitqkqhysesadptais 281
QY 500 PRDRKHPRVMTLPPCPPDMDLMEAKDQVAFELMRTFKLPCEKINDWVPYDRDE 559
DB 282 gmkrhdsrvidfpdptndlmieakeqavfelcrryel-----q 325
QY 560 NRPAPPVKAPKKKGKRRKRTTDEEAPEEVDTAADDVKAPEGPKEVPEERAMGPGY 619
DB 326 nppcp-----leimgpe-----yqdtldg----- 344
QY 620 NRVVWPLGCEEWLKPKEVKKGKVPVEVEDE 651
DB 345 ---yypgaekrltarkrrsrk-----eeveed 369

RESULT 7
AAY44505
ID AAY44505 standard; Protein; 294 AA.
XX
XX AAY44505;
XX
XX 27-MAR-2000 (first entry)
XX
XX D. radiodurans delta228-UV damage endonuclease.
XX
XX Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;

KW glutathione-S-transferase signal peptide; uvrl+ gene product;
 KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
 KW abasic site; aplatinum diaduct; mismatched nucleotide pairing;
 KW nucleotide alkylation; skin cancer.
 XX
 OS Dienococcus radiodurans.
 PN WO9963828-A1.
 XX
 XX 16-DEC-1999.
 XX
 XX 08-JUN-1999; 99WO-US12910.
 XX
 XX 08-JUN-1998; 98US-0088521.
 PR
 PR 18-MAY-1999; 99US-0134752.
 XX
 XX (UYEM-) UNIV EMORY.
 PA
 XX Doetsch PW, Kaur B, Avery AM;
 PI
 XX WPI; 2000-116417/10.
 DR
 XX A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers -
 PT
 PS Claim 16; Page 60; 133pp; English.
 XX
 XX The present sequence is delta228-UV damage endonuclease from
 CC D. radiodurans. Delta228-UVDE is a 288 amino acid deletion of the
 CC N-terminal of the uvrl+ gene product. This is expressed in frame with a
 CC GST leader sequence to generate a fusion protein. This provides stable
 CC endonuclease fragments for cleaving a double-stranded DNA molecule that
 CC has a distorted structure resulting from UV radiation damage, a
 CC photoproduct, an abasic site, mismatched nucleotide pairing, aplatinum
 CC diaduct, an intercalated molecule or alkylation of a nucleotide. Uvrlp
 CC can be used in compositions for internal or topical application and as a
 CC therapeutic agent for skin cancers.
 XX
 SQ Sequence 294 AA;

 Query Match 7.5%; Score 262.5; DB 21; Length 294;
 Best Local Similarity 32.2%; Pred. No. 1.1e-14;
 Matches 85; Conservative 38; Mismatches 102; Indels 39; Gaps 13;

 QY 292 YGIRFLRLSSEFP---FASHPVHGKYLAPFASEVLAEAGRAVAELGRLTHPTPCQFTQL 348
 DB 56 hdirlyrlssifpmlldagddtgaavithlapqil-eaghaftdagvrlmhpqfivl 114

 QY 349 GSPRKEVESATRDLEYHDEL---LSILKLPEQQNRDVAVLIH-MGQFCGDKAATLERF 403
 DB 115 nsdrpevressvramsharvmdgigltartpwn-----lllhgkggrgaelaali--- 166

 QY 404 KRNARLSQSCNRLVLENDVDGVTVDHLLPYCEELNIPVMVDYHHNHCDFPAHLREGT 463
 DB 167 ---pdlpdpvrlrlglenderayspaelliceatgtplvfadahhvv-----hdkl 215

 QY 464 LDISPDKLOE---RANTWKRKGIKQKMHYSEPCDCAVTPRDRKRHRPVMTLPCCPDM 520
 DB 216 pqgedpsvrewlratatqppe-wqvvhslnglieg---pqdirr-hshliadffsayadv 270

 QY 521 -DLMIEAKDKQAV--FELMRTFK 541
 DB 271 pqieveakqkeaaialrlmapfk 294

 RESULT 8
 ID AAY85657
 AC AAY85657 standard; Protein; 1341 AA.
 XX
 AC AAY85657;
 XX
 DT 12-FEB-2001 (first entry)

XX Human Acinus L protein sequence.
 DE
 XX Chromatin aggregation induction; human; Acinus; AIDS; herpes; diabetes;
 KW Alzheimer's disease; Parkinson's disease; cardiac infarction;
 KW brain infarction; adenovirus infection; viral hepatitis.
 XX
 XX Homo sapiens.
 OS
 XX WO2000061743-A1.
 PN
 XX 19-OCT-2000.
 XX
 XX 07-APR-2000; 2000WO-JP02254.
 PR
 XX 09-APR-1999; 99JP-0103317.
 PR
 XX (SHIO) SHIONOGI & CO LTD.
 PA
 XX Sahara S, Eguchi Y, Tsujimoto Y;
 PI
 XX WPI; 2000-665130/64.
 DR N-PSDB; AAC61196.
 DR
 XX A novel polypeptide which induces chromatin aggregation but is not
 PT involved in DNA fragmentation used for the regulation of apoptosis and
 PT screening of potential apoptosis inhibitors -
 PT
 PS Disclosure; Page 60-67; 99pp; Japanese.
 XX
 XX This invention relates to a polypeptide which induces chromatin
 CC aggregation. The polypeptide is a fragment of the human Acinus protein.
 CC included in the invention is a nucleotide sequence encoding the
 CC polypeptide, antisense oligonucleotides, antibodies recognising the
 CC polypeptide sequence, and apoptosis regulating agents containing the
 CC polypeptide, nucleotide sequence, and sense or antisense
 CC oligonucleotides. Also included in the invention is a method for
 CC screening compounds for their activity as chromatin aggregation
 CC inhibitors, regulators or promoters. The polypeptide induces chromatin
 CC aggregation within the cell nucleus without inducing DNA fragmentation.
 CC The polypeptide can be used in the investigation and treatment of
 CC diseases in which apoptosis is implicated, including AIDS, Alzheimer's
 CC disease, Parkinson's disease, cardiac or brain infarction, herpes or
 CC adenovirus infection, diabetes and viral hepatitis. The present sequence
 CC represents the human Acinus protein.
 XX
 SQ Sequence 1341 AA;

Query Match 4.6%; Score 160.5; DB 21; Length 1341;
 Best Local Similarity 19.6%; Pred. No. 8.8e-05;
 Matches 153; Conservative 104; Mismatches 315; Indels 209; Gaps 29;

 QY 1 MPKRKSKAAALDTPQSESTFSTLSSAPSPARNLRRSGRNILQPSSEKDRD-----H 54
 DB 565 lpsrlrkqsa---dsstsrsssssssssrssrpsdssgrrshslrakqrdaqrth 620

 QY 55 EKRSGEELAG-----RMWGDANGHCLR-----EGKEDEEGVKM 88
 DB 621 anprgrpkmdsrstseersrstrssasnsrkslspgvsrdsstsyetdkpssgqev 680

 QY 89 ATEGLAHM-----ERRLQKATKROKKQLEEDGIPVPSVVSRRPPTAPYHHKSTNAERE- 141
 DB 681 alppvpqlqvcepkertststsssvqarrlrsqpesaekhvtqlrlqpergspkceaeap 740

 QY 142 --AKEP-----VLKTHSKD-----VERAEICVD-----DVVKMEPAATNIEPED 180
 DB 741 paatqpqlsetqshlpeserlihtveekcvmdtsenrpendv----pepmpiadqv 796

 QY 181 AQDAARERGAARPFAVNSSYLPLPWKGRLE-----GYACINTLYLRNAKP----- 222
 DB 797 sndrpegsvdeekesslpksfkrkksivsvsatkgvpagnsdteggqgqrkrwgasta 856

QY 223 -----PIFSRTCRMASIVDHRHLPQDEPEHHLKNNKPKDSKEPODEL---GHKFVQEL 274
 Db 857 ttqkpsittesklslpdkplaag---eavvdilhaddrsisedeternddgthdk 913
 QY 275 GLANARDIVKMLCWNKYGIRFLSLSEMPFFASHPVHGYKLAPFASEVLAEAGRAVAEL 334
 Db 914 glkicrtvtgvpagenggreceeekepeaepvpqvsvevalpppaehevkvvtl 973
 QY 335 GHLRTH-----PCQFTQLGSPRKEVVEASIRDEYHDELL---SLLKLP 376
 Db 974 gdtlrrsisqksgvstidddpvtatqvsppr---gkislvlhslnlvrftlgql 1029
 QY 377 EQQRADAVMIITHGGQFGDKAATLERKRN---YARLSQSKNRLVLENDVDGWTVDHL 432
 Db 1030 ellqrtgtlv-----eeafwidkikshcftystveaavrtalhg--vkwpqsnp 1079
 QY 433 LPVCEELNIPWLDYHHNTCFDPAHLREGLTLDISDPKLOERTANTWKRGIKQKMYSE 492
 Db 1080 kflcadvaedeldyhrglivdrp-----setktee-----ggiprplhppp 1121
 QY 493 PCDGAVTPDRRRKRPVMTLPPCPDMDLMIKAKDEQAVFE--LMRTFKLPGEKIND 550
 Db 1122 p-----ppvqpqhpraegreavreqwaereremerrtrrs 1161
 QY 551 MNPYDRD-----DENRAPPVPKPKKKGKRRTTDEEAAEPVEVDTAADDV---KDA 601
 Db 1162 erewdrkvregprsrdrtrrkerakskekkekaqepakliddlfrtkaa 1221
 QY 602 PEGRKEPVEERANGPNRYVW-PL-----GCEWMLKPKRKVKKGVPEVEDEG 652
 Db 1222 p-----ciwlpitdsqivqkeaaeraerakererkegeeeeqker 1263
 QY 653 E 653
 Db 1264 e 1264
 RESULT 9
 AAB94420
 ID AAB94420 standard; Protein: 914 AA.
 AC AAB94420:
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:15022.
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX Homo sapiens.
 OS
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602
 full-length cDNAs defined in the specification, and for the detection
 and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -
 XX
 PS Claim 8: SEQ ID 15022; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 full-length cDNAs defined in the specification. Where a primer set
 comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 to the complementary strand of a polynucleotide which comprises one of
 the 5602 nucleotide sequences defined in the specification, where the
 oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 of an oligonucleotide comprising a sequence complementary to the
 complementary strand of a polynucleotide which comprises a 5'-end
 sequence and an oligonucleotide comprising a sequence complementary to a
 polynucleotide which comprises a 3'-end sequence, where the
 oligonucleotide comprises at least 15 nucleotides and the combination of
 the 5'-end sequence/3'-end sequence is selected from those defined in
 the specification. The primer sets can be used in antisense therapy and
 in gene therapy. The primers are useful for synthesizing polynucleotides,
 particularly full-length cDNAs. The primers are also useful for the
 detection and/or diagnosis of the abnormality of the proteins encoded by
 the full-length cDNAs. The primers allow obtaining of the full-length
 cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 represent oligonucleotides, all of which are used in the exemplification
 of the present invention.
 XX
 SQ Sequence 914 AA;

Query Match 4.1%; Score 143.5; DB 22; Length 914;
 Best Local Similarity 19.3%; Pred. No. 0.0016;
 Matches 156; Conservative 103; Mismatches 281; Indels 269; Gaps 39;

QY 11 LDTPOSESTF-----SSTLDSSAPSPARNLRSGRNILQPSSEKDR-DHEKRSGEE-LA 63
 Db 100 vdpllkstfdenpseteeaaagigklegedgvd-kclsekdytdtsidsleenld 158
 QY 64 GRMMCK---DANGCHLRECKELEGVKMAI-----EGLARMERLQR--ATKROKKOLE 112
 Db 159 kkkkgkfkpeasdrclrsqldssadrcrlrqssdsasactelkvpknpsakrskkegh 218
 QY 113 EDGIPVPSVVSFRPTAPYHHKSTNAEREAKEPVLKTHSKDVEREAIEGVDVVMPEAA 172
 Db 219 pygtitpklgll-----pdsfhtetletlekpvsnerpsekdaeegegg--giitrtql 269
 QY 173 TNIIEPDQDAAGAAARPPAVNSSYPLPWKRLGYACLNTYLRNAKPPIFSSTRCRM 232
 Db 270 kmldkevkgrlgeifpsrdp-ittagqplpgerleiyvqskmdeknahip-----s 320
 QY 233 ASIVDHRHPLQFEDPEH-----HLKNNPKSKPEQDELGHKFVQELGL 276
 Db 321 eslackrdpeakeepghiptqhveeavnevndentqkddesdapsalg-----lss 374
 QY 277 ANARDIVKMLCWNKYGIRFLSLSEMPFFASHPVHGYKLAPF-ASEVLAEAGRAVAELG 335
 Db 375 ssgsdaara---pksvprpkrltstynlr---hahslgldaskvuse--keaaqvn 424
 QY 336 HRLTHPGQFTQLGSPRKE-----VVESAIRDEYHDELLSLKLPEQQRNDAVMIHMG 390
 Db 425 pmpkenda-sesgdpldeddvtdvvdqepkfmew-----caeeenqe--lianfn 472
 QY 391 GGF-----GDKAATLERFKRNYARLSQSKNRLVLENDVDGWM 427
 Db 473 adymkvqkwqiwlekegqptprarnksdklkeiwskskr---srkcrsslesqk----- 523
 QY 428 TVHDLILPVCEELNIPWLDYHHNTC-----
 Db 524 -----cspvqmlfmfnklsnvckwlflettrslvkvklntrlpvggvpvkhplq 575
 QY 454 -EDPAHLREGTLTIDSPKLQERI-----ANTWKRGIKQKMYHS-----EPCDGA-V 498
 Db 576 kyapsslypssliqae--rlkkhllkfpqatpaknnwkmqklwakfrenpdqvepedsdv 633

QY 103 ATKROKQLEEDGIPVSVSREPTAPVHHKSTNAEEREKEPV----- 146
 Db 2547 rtrkekaraleldvofkelfs----apgteesntdkntkipcksppltdtatctkr 2602
 QY 147 -----LKTHSKDVERAEI-GVDDVVRMEPAA-----TNII-----EPEDAQDAAE 186
 Db 2603 cpktrprkeveelsaverlqtsgqstthkpeasgdegikvikgrakkkpnpveeeps 2662
 QY 187 RCAARPPAVNSYLLPKNKGRGLVACLTLYLRNAKPPITFSSRTCRMASIVDHRPLOFED 246
 Db 2663 rrrprapkeaq---plidlagltelsetsghtagkatk---ipcespplvvd 2715
 QY 247 ---EPEHHLKKNPKDSKPEQDELGHKFVQELGLANARDIVKMLCWNKYGIRFLRLSSEM 303
 Db 2716 ttastkrhlrtvkvqveepsavkftqsg--ettidakepatedk-gikalakesakq 2772
 QY 304 PPFASHPVHGYKLAPFASEVLAE-----AGRYAAELGH-----RLTHP-----QOFT 346
 Db 2773 tpapaasvtgsrrrrprapresaqaidlagfkdpaghteesmtdtktkipcksspele 2832
 QY 347 QLGSPRKVEVESAIRDLEVHDELLSLKL-----PQQN 380
 Db 2833 dtatskrirpratrackvekeellavglktqstgetthtdkepvgeqgktafkqpakrn 2892
 QY 381 RDVAMIHWGGO---FGDKAATLERFKRNYARLSQSKNRLVLENDVGVTVHDLPLVCE 437
 Db 2893 vdaedvigsrrprapkeagpld--asfqlsq----- 2927
 QY 438 ELNIPWLDYHHNICFPDAHLRE-----GTLDISDP-KLOBRIANTWKRK 482
 Db 2928 -----pghteelangaadsftsapktqtdsgkplkistrvirap-- 2967
 QY 483 GIKOKMHYSEPCDGAVTPRDRKRHRPVMT-LPPCPDMDMLIEAKDKEQAVFELMRTFK 541
 Db 2968 -----vepygdvsvtrdpvksqskntsllpplp-----fkgggkgdgsvtgktrlc 3014
 QY 542 LPEFEKINDVPYDRDENRAPPVKKPKKKGGKRRKTTDEEAPEEVEDTAADVDKDA 601
 Db 3015 mpapeelvelpasakkrqvrprargkssepyvkmksrltsakrlpeae-einsndmktm 3073
 QY 602 PRGPK---EVPEE 612
 Db 3074 keehklqdsvenk 3087

RESULT 13
 AAM14533
 ID AAM14533 standard; Protein; 2665 AA.
 XX
 AC AAM14533;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Peptide #967 encoded by probe for measuring cervical gene expression.
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200157278-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00670.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

XX
 PA
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488901/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 PS Claim 27: SEQ ID No 19359; 487pp; English.
 XX
 CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs: see AAI10068-AA128459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2665 AA:
 Query Match 3.9%; Score 136; DB 22; Length 2665;
 Best Local Similarity 18.8%; Pred. No. 0.034;
 Matches 147; Conservative 95; Mismatches 296; Indels 242; Gaps 31;
 QY 3 SRKSKAAALDTPOSESTFSSTLD-----SSAPSP-----ARNLRSGCNILQPSSEKD 51
 Db 840 sreenwslfd-dsrfanfrnkdkkvsapripipswymkkkkirtdsegkmdkkeh 898
 QY 52 RDHEKRSGEELAGRM-----GKDN--GHCLRE-----GKEQEGVYKMAIEG 92
 Db 899 keeegerqelfasrlfshssifedqskrlghlerkeedsdfisrgiykqtseanstds 958
 QY 93 LAR-----MERLORATYKROKKQLEEDGIPVSVSREPTAPVHHKSTNA-----EEREAKE 144
 Db 959 iqepvvlfrsfmeltmrmqdeked--qkpkveqkedtenhpktpesapenkdselkt 1016
 QY 145 p-----VLKTHSKDVERAEIGVDDVVVKMEPAATNIEPEDAQDAAE 186
 Db 1017 ppsvgpsvtvtvlesapsalekttdgtktveapl-vteektvepatsv---eeakpase 1071
 QY 187 RGAARPPAVNSYLLP-----LPWKGRUGYACLTLYLRNAKPPITFSSRTCRMASI 235
 Db 1072 papapvseqleqvdllpgadpdkcaammpagveegsgdppylidakpptpgasfsqaen 1131
 QY 236 VDHR-----HPLOFEDEPEHHLKKNPKDSKEPQDELGHKFVQELGLANARDIVKMLC 287
 Db 1132 vdpedstqplskpakkecanepkaekpdadadapdangk----- 1173
 QY 288 WNEKYGIRFLRLSSEMPFPASH-----PVHGYKLAPFASEVL 324
 Db 1174 -----aaapesqpasedleivdpvaakdkkpnkskrsktpvqaaavsiyekpvt 1224
 QY 325 AEACRYAAELCHRLTTHPGQFTQLGSPRKEVVESEATRDLEYHDELLSLKLPSQONRDV 384
 Db 1225 rkserldre-----klkrsnprgea-----qkllkmeakekirtas 1263
 QY 385 MITHMGQFGDKAATLERFKRN-----YARLSQSKNRLVLENDVGVTVHDLPLVCEEL 439
 Db 1264 knsaadliepelpsllsrtrrvsvyatungdh-eur-----spvkepv 1307
 QY 440 NIPW-----LDYHHHNICFPDAHLRECTLDISDPKLOERTANTWKRKGKQKMHYSEPCD 495
 Db 1308 eqprvtrkrlereleqaaavpttprg-----rpktrrr-adeeeneakepaetlkppe 1362
 QY 496 GAVTPDRRR-----KHRRPV-MTLPCCPMDMLIEAKD-----KQOAVF 534
 Db 496 GAVTPDRRR-----KHRRPV-MTLPCCPMDMLIEAKD-----KQOAVF 534

(MOLE-) MOLECULAR DYNAMICS INC.

Db 1363 gwrsprsqktaaggpgqkgkgnepkvdatrpeatttevgqigvksessmepkaaeaeags 1422
 QY 535 ELMRTFKLPGEKINDMVPYDRDDENRPAVPKPKKKG----- 574
 Db 1423 eqkrdkdgatknppetavvevkkpape-knskskrgrsrnsrlavdksaslknvda 1481
 QY 575 -----GRRK-----RTTDEAAAEPEVDTAADVKDAPEGPKVEPE-EERAMGGP 618
 Db 1482 avsprgaagaqageregsvavspkespqkedglssqklsdpvdpdkepekedvsasgp 1541

RESULT 14
 AAM26950
 ID AAM26950 standard; Protein; 2665 AA.
 XX
 AC AAM26950;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Peptide #987 encoded by probe for measuring placental gene expression.
 XX
 KW Probe: microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN W0200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLF-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 XX WPI; 2001-488897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human placenta -
 XX
 PS Claim 27; SEQ ID No 27219; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENPs;
 CC see AA131315-AA157546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.
 XX
 SQ Sequence 2665 AA;

Query Match 3.9%; Score 136; DB 22; Length 2665;
 Best Local Similarity 18.8%; Pred. No. 0.034;
 Matches 147; Conservative 95; Mismatches 296; Indels 242; Gaps 31;
 QY 3 SRKSKAAALDTQSSSTSSLTLD-----SSAPSP-----ARNLRRGRNLLQPSSEKD 51
 Db 840 screenswfdw-dsrfanfrnknkdvdsapripwmykknkirtsdsegkmdkdh 898
 QY 52 ROHEKRSGEELAGRMW-----GKDNAN--GHCLURE-----GKFOEGYKWMATEG 92
 Db 899 keeqqerqelfasrlhssifedqskriqlhlerkeedsdfisgriyqktsegansttds 958

QY 93 LAR-----MERRLQRATKQKQLEEDGIPVSVVSRFTAPYHHKSTNA-----BEREAKE 144
 Db 959 lqepvvlfhsrfmeltrmqkqekkd--qkpkvevkedqtenhpktpesapenkdselkt 1016
 QY 145 P-----VLKTHSKDVEREAEIGVDVVKMEPAATNIIEPEDAODAAE 186
 Db 1017 ppsvgppsvvtlesapsalekttgdkvtveapl-vteektvepatsv-----eeakpase 1071
 QY 187 RCAARPPAVNSYLP-----LPWKGRGLYACUUNTYLRNAKPPIFSSRTRCMAEI 235
 Db 1072 papapveqlqevdlppgaadpkkaeammpagveegssgqppyldakpptgpaasfsqaesn 1131
 QY 236 VDHR-----HPLQFDEDEPHHLKNPKDKSKPQDELGHKFVQELGLANARDIVKMLC 287
 Db 1132 vdepdstqplskpaqkseanepkaekpdacadaepdangk----- 1173
 QY 288 WNEKYGIRFLRISSEMPFPASH-----PVHGYKLAPFASBVL 324
 Db 1174 -----aaapesqpasedievppvaakdkkpnkskrsktpvqaaavsiavekpyt 1224
 QY 325 AEAGRVAAELGHLRTHPGQFTQGLSPKKEVVESAIRDELYHDELLSLKLPLEQONRDVA 384
 Db 1225 rkseridre-----klksusprgea-----qkllkmeaekitrtas 1263
 QY 385 MIIHMGOGFGDKAATLERKRN-----YARLSQCKNRLVLENDVGVTVHDLPLVCEEL 439
 Db 1264 knsaadlehepsiplsrtrrnrvsvyatmgdh-enr-----spvkepv 1307
 QY 440 NIPMV-----LDYHHNICFDPAHLREGTLDISDPKLOERTANTWKRKGIKOKMHYSEPCD 495
 Db 1308 eqprvtrkrlereiqeaaavptprg---rrpktrrr-adeeeeneakepaetllkpe 1362
 QY 496 GAVTPRRRR-----KHRPRV-MTLPPCPPDDMLMIEAKD-----KEQAVF 534
 Db 1363 gwrsprsqktaaggpgqkgkgnepkvdatrpeatttevgqigvksessmepkaaeaeags 1422
 QY 535 ELMRTFKLPGEKINDMVPYDRDDENRPAVPKPKKKG----- 574
 Db 1423 eqkrdkdgatknppetavvevkkpape-knskskrgrsrnsrlavdksaslknvda 1481
 QY 575 -----GRRK-----RTTDEAAAEPEVDTAADVKDAPEGPKVEPE-EERAMGGP 618
 Db 1482 avsprgaagaqageregsvavspkespqkedglssqklsdpvdpdkepekedvsasgp 1541

RESULT 15
 AAM02259
 ID AAM02259 standard; Protein; 2665 AA.
 XX
 AC AAM02259;
 XX
 DT 09-OCT-2001 (first entry)
 XX
 DE Peptide #941 encoded by probe for measuring human breast gene expression.
 XX
 KW Probe; human; breast disease; breast cancer; development disorder;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 OS Homo sapiens.
 XX
 PN W0200157270-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 29-JAN-2001; 2001WO-US00661.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast
XX
XX Claim 27; SEQ ID No 10999; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes
XX (see AA10010-A110067). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for measuring human gene expression in
XX a human breast sample, where the probe hybridizes at high stringency to a
XX nucleic acid expressed in the human breast. The probes are useful for
XX predicting, diagnosing, grading, staging, monitoring and prognosing
XX diseases of the human breast, particularly those diseases with polygenic
XX aetiology. The diseases include: breast cancer, disorders of development,
XX inflammatory diseases of the breast, fibrocystic changes, proliferative
XX breast disease and non-carcinoma tumours.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2665 AA;

Query Match 3.9%; Score 136; DB 22; Length 2665;
Best Local Similarity 18.8%; Pred. No. 0.034;
Matches 147; Conservative 95; Mismatches 296; Indels 242; Gaps 31;
QY 3 SRKSKAALDTQSESSSTFSSFLD---SSAPSP-----ARNLRRSRGNLQPSSEKD 51
DB 840 sreenwsfldw-dsrfanfrnnkkekvdssapripwymmkkkirtsegmddkkedh 898
QY 52 RDHEKRSGEELAGRMW-----GKDAN--GHCLRE-----GREQEEGVKMAIEG 92
DB 899 keeqeqelfasrfhsifeqsdskriqhlrkeedsdfisgriyqktsegansttds 958
QY 93 LAR---MERRLORATKQKOLEEDGIPVSVVSRRFTAPYHHKSTNA---BEREAKE 144
DB 959 idqepvvlfsfrfmeltrmqgkekd--qkpevekedtenhpktpesapenkdselkt 1016
QY 145 P-----VLKTHSKDVEREAEIGVDVVRKMEPAATNIIIPEDAQDAAE 186
DB 1017 ppsvppsvvtvlesapsalekttgdkttveapl-vteektvepatvs---eeakpase 1071
QY 187 RGAARPPAVNSSYLP-----LPWKGRGLGYACLTNLYLRNAKPPIFSSRTRCMASI 235
DB 1072 papapveqlqevdlppgagdkpkeaaampagveegssgdpqpyldakpbtptgafsfqaesn 1131
QY 236 VDHR-----HPLQFDEPEHHLKNPKSKPEQDELGHKFVQELGLANARDIVKMLC 287
DB 1132 vdpedstqplskpaqkseeaneapkaekpdatadaepdanq-----1173
QY 288 WNEKYGIRFLRLSSEMFPPASH-----PVHGYKLAPFASEVL 324
DB 1174 -----aeapesqppasedlewdppvaakdkkpnkskrsktpvgaavaavivekpvt 1224
QY 325 AEAGRVAAELGHLRTHPGQFTQLGSPKREVVEATRDLEYHDELLSLKLPLEQONRDV 384
DB 1225 rksridre-----klkrnsprgea-----qklllekmaekitrtas 1263
QY 385 MITHMGQGFQDKAATLERKRN-----YARLSQSKCNRLVLENDVDVGTVHDLPLVCEEL 439
DB 1264 knsaadlehepslplsrrrrnrvsvyatmgdh-enr-----spvkepv 1307
QY 440 NTPMV-----LDYHHNHCIFDPAHLREGTLDISDPKLOERIANTWKRRGKIKQKMHYSEPCD 495
DB 1308 eqprvtrklrelgeaavaavtpprrg-----rppktrrr-adeeeeneakepaetlkppe 1362

QY 496 GAVTPRDRS-----KRRPRV-MTLPPCPDMDLMEAKD-----KEQAVF 534
DB 1363 gwrfsrsqktaagggpggkkgknepkvdatrpeattevgpqigvkessmepkaaeags 1422
QY 535 ELMRTFKLFCFEKINDMVPYDRDDENRRPAPPYKAPKKKKG-----574
DB 1423 eqkrdr:kdsgrtdknppetapvevvekkpape-knskkskrgrsrnslavdksasiknvda 1481
QY 575 -----GRRK-----RTTDEEAAEPVEVDTAADDVKDAPEGPKVEPE-EERAMGCP 618
DB 1482 avsprgaaagaderesgvvavspkespqkedglsqiksdqvpdpdkepekedvsasgp 1541

Search completed: January 15, 2002, 13:50:07
Job time: 296 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:52:23 ; Search time 74.94 Seconds
(without alignments)
666.807 Million cell updates/sec

Title: US-09-724-296-36
Perfect score: 3482
Sequence: 1 MPSRKSAAALDTPQSEST.....REVKKGVPEVEDEGEFDG 656

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR68.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3472	99.7	656	2 S55262	UV-endonuclease -
2	1011.5	29.0	599	2 S71134	UV-endonuclease -
3	377	10.8	322	2 E83974	hypothetical prote
4	348.5	10.0	320	2 S55418	UV-endonuclease ho
5	261.5	7.5	326	2 C75350	probable UV damage
6	160.5	4.6	1280	2 T00365	hypothetical prote
7	155.5	4.5	5327	2 T13564	microtubule-associ
8	146	4.2	796	2 S56231	hypothetical prote
9	141.5	4.1	1110	2 I51116	NF-180 - sea lamp
10	140	4.0	727	2 T17292	hypothetical prote
11	139.5	4.0	1510	2 T33100	hypothetical prote
12	139	4.0	1815	2 C81169	IgA-specific metal
13	138	4.0	1547	2 T28657	blackjack protein,
14	137	3.9	1560	2 T42727	cell proliferation
15	137	3.9	2897	2 B48666	cell proliferation
16	137	3.9	3256	2 A48666	cell proliferation
17	136.5	3.9	845	2 A45669	neurofilament trip
18	136.5	3.9	849	2 S00030	neurofilament trip
19	136	3.9	3942	2 T42730	Bassoon protein -
20	135.5	3.9	971	2 T24866	hypothetical prote
21	135	3.9	793	1 JH0628	caldesmon - human
22	134	3.8	2663	1 S28261	centromere protein
23	133.5	3.8	980	2 E71606	hypothetical prote
24	132.5	3.8	915	2 S54485	CES1 protein - yea
25	132.5	3.8	6642	2 T29757	protein UNC-89 - C
26	132	3.8	734	2 F85073	hypothetical prote
27	131.5	3.8	1222	2 T24490	hypothetical prote
28	131	3.8	606	2 A43427	neurofilament trip
29	128.5	3.7	856	2 T16543	hypothetical prote

30	128.5	3.7	7962	2 I38346	elastic titin - hu
31	128	3.7	1020	1 QFHHH	neurofilament trip
32	128	3.7	1805	2 A34736	nestin - rat
33	128	3.7	3488	2 T34418	hypothetical prote
34	128	3.7	4151	2 T13734	groovin gene prote
35	127.5	3.7	2453	2 S60254	nuclear receptor c
36	126.5	3.6	720	2 T26819	hypothetical prote
37	126.5	3.6	2416	2 T13825	adenomatous polyo
38	126	3.6	522	2 C96608	hypothetical prote
39	126	3.6	1773	2 A81937	IgA-specific metal
40	125.5	3.6	786	2 C86406	hypothetical prote
41	125.5	3.6	1400	2 T52359	hypothetical prote
42	125.5	3.6	2441	2 D71623	erythrocyte membra
43	124.5	3.6	1829	2 T76135	hypothetical prote
44	124.5	3.6	4377	2 A55575	ankyrin 3, long sp
45	124	3.6	1072	1 A37221	neurofilament trip

ALIGNMENTS

RESULT 1

S55262

UV-endonuclease - Neurospora crassa

C:Species: Neurospora crassa

C>Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000

C:Accession: S55262

R:Yajima, H.; Takao, M.; Yasuhira, S.; Zhao, J.H.; Ishii, C.; Inoue, H.; Yasui, A.

EMBO J. 14, 2393-2399, 1995

A:Title: A eukaryotic gene encoding an endonuclease that specifically repairs DNA dam

A:Reference number: S55262; MUID:95292980

A:Accession: S55262

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-656 <YAJ>

A:Cross-references: EMBL:D11392; NID:q1526560; PID:9927215

C:Genetics:

A:Gene: uvel

Query Match	99.7%	Score 3472:	DB 2:	Length 656:
Best Local Similarity	99.7%	Pred. No. 2.2e-208:		
Matches 654:	Conservative 1:	Mismatches 1:	Indels 0:	Gaps 0:
QY	1	MPSRKSAAALDTPQSESTFSSTLDSASAPARNLRSGRNILQPSSEKDRDHRKRSGE	60	
Db	1	MPSRKSAAALDTPQSESTFSSTLDSASAPARNLRSGRNILQPSSEKDRDHRKRSGE	60	
QY	61	ELAGRMGKDANGHCLREGKQEGEGVKMAIEGLARMERRIQATKQKKOLEEDGIPVPS	120	
Db	61	ELAGRMGKDANGHCLREGKQEGEGVKMAIEGLARMERRIQATKQKKOLEEDGIPVPS	120	
QY	121	VVSFPTAPYHHKSTNAEREAKPEVLKTHSKDVERAEIGVDVVVKMEPAATNIEPED	180	
Db	121	VVSFPTAPYHHKSTNAEREAKPEVLKTHSKDVERAEIGVDVVVKMEPAATNIEPED	180	
QY	181	AODAAERGAARPPAVNSSYLPLPMKGRGLGYACLTNTYLNRNAKPPIFSSRTRCMASIVDHRH	240	
Db	181	AODAAERGAARPPAVNSSYLPLPMKGRGLGYACLTNTYLNRNAKPPIFSSRTRCMASIVDHRH	240	
QY	241	PLQFEDEPEHLLKNPKSKPEODELGHKFEVDELGLANARDIVKMLCWNEKYGIRFLRLS	300	
Db	241	PLQFEDEPEHLLKNPKSKPEODELGHKFEVDELGLANARDIVKMLCWNEKYGIRFLRLS	300	
QY	301	SEMPFFASHPVHGYKLAPFASEVLAEGRAVAEEIGHRLTTHPQFTQLGSPRKEVESA	360	
Db	301	SEMPFFASHPVHGYKLAPFASEVLAEGRAVAEEIGHRLTTHPQFTQLGSPRKEVESA	360	
QY	361	RDLEYHDELLSLKLPQONRODAVMIHHMGQFGDKAATLFRFKRYARLSQSCKNRVL	420	
Db	361	RDLEYHDELLSLKLPQONRODAVMIHHMGQFGDKAATLFRFKRYARLSQSCKNRVL	420	
QY	421	ENDDVGTWHDLLPVCEELNIPMVLDYHHHNTCFDPAHLREGTLDISDPKQLOERIANTWK	480	

```

Db 421 ENDDVGVTHDLPVCELNIPVLDYHHHNCFCFPAHLREGTLDISDPKLOQRANTWK 480
QY 481 RGIKQKMHYSEPCDGAATPRDRKRRPRVMTLPCCPDMDLMEAKDKEQAVFELMRTF 540
Db 481 RGIKQKMHYSEPCDGAATPRDRKRRPRVMTLPCCPDMDLMEAKDKEQAVFELMRTF 540
QY 541 KLPGFECTNDVYDRDDENRPPVVPKAPKKKGKGRKRTTDEEAAPPEEVDVTAADDVKD 600
Db 541 KLPGFECTNDVYDRDDENRPPVVPKAPKKKGKGRKRTTDEEAAPPEEVDVTAADDVKD 600
QY 601 APEGKPEVPEERAMGGPNRYVWPLGCEEWLKPKEVKKVGGKVPVEDEGEFDDG 656
Db 601 APEGKPEVPEERAMGGPNRYVWPLGCEEWLKPKEVKKVGGKVPVEDEGEFDDG 656

RESULT 2
UV-endonuclease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 20-Jun-2000
C:Accession: S71134; T39815
R:Takao, M.; Yonemasu, R.; Yamamoto, K.; Yasui, A.
Nucleic Acids Res. 24: 1267-1271, 1996
A:Title: Characterization of a UV endonuclease gene from the fission yeast Schizosaccharomyces pombe
A:Reference number: S71134; MUID:96188860
A:Accession: S71134
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-599 <TAK>
A:Cross-references: EMBL:D78571; NID:gi399000; PIDN:BAAL1415.1; PID:gi399001
A:Experimental source: strain S972
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Lucas, M.; Gaillardin, C.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z21881
A:Accession: T39815
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-599 <LYN>
A:Cross-references: EMBL:AL023859; PIDN:CAA19577.1; GSPDB:GN00067; SPDB:SPBC19C7.09c
A:Experimental source: strain 972h; cosmid cl9C7
C:Genetics:
A:Gene: UVDE; SPBC19C7.09c
A:Map position: 2

Query Match 29.0%; Score 1011.5; DB 2; Length 599;
Best Local Similarity 36.6%; Pred. No. 1.9e-55;
Matches 241; Conservative 88; Mismatches 149; Indels 181; Gaps 18;

QY 3 SRKKAALDTPQSESTFSTLSDSAPSARNLRSGRNILQPSKDRDHRKSGBEL 62
Db 110 ARSKKATDSVSKIDESVASYDSST-----HLRRSSR-----SKKPNYNSSESE- 157
QY 63 AGRMGKDANGCHLUREGKEGEGVMAIEGLARMERLORATKROKOLEED----- 114
Db 158 -----SEQISKATKKVKKKEEVEEVEVDEK 184
QY 115 GIPVPSVVSRE-PTAPYHHKSTNAEAEKPEVLKTHSKDVERAEIGVDDVVKMEPAAT 173
Db 185 SLKNSSDEEPVVPVPEOLETPISKRRS-----RSSAKNLEKESTMNLDDHA----- 232
QY 174 NIIEPEDAQAARGAARPPAVNSYLLPLPWKGLGYACLTNTYLRNAKPPIPSSRTCRMA 233
Db 233 -----PREMFCDLK-----PIWRGLGYACLTNTYLRNAKPPIPSSRTCRIT 275
QY 234 LIVDRHPLQFEDRPEHHLKNKPKSKPEQDELCHKFQVQELGLANARDIVKMLCWNEYK 293
Db 276 TI-----QRGLESVKQLGTQNVLDLILKLVNHNHNEG 307
QY 294 IRFLRLSEMFPFASHPVHGYKLAFFASEVIAEAGRAVAELGHLRTHPCQFTQLGSPRK 353
Db 308 IHFRVSSDLFPFASHAKYGYTL-EFAQSHLEEVGKLANRYNHLRTHMPCQYQTQIASPRE 366

```

```

QY 354 EVESAI RDLLEYHDELLSLKLPQGNRDVAMLIHMGQFQDKAATLERKKNYARLSQS 413
Db 367 VVDSAI RDLAYHDEILSRWKLNEQKQAVLIHLGCTFEGGKETLDRKKNYQRLSDS 426
QY 414 CKNR[AVLENDVGVTHDLPVCEELN(PMVLGYHHHNTCFDPAHLREGTLDISDPKLOE 473
Db 427 VKARLVLENDVSVQDLPLCQELNPLVLDWHHHNTV--PGTLREGSLDLM--PLIP 482
QY 474 RIANTWRKGTGKQMHYSEPCD-GAVTPDRRRKRRPRVMTLPCCPDMDLMEAKDKEQA 532
Db 483 TIRETWIRKGTGKQMHYSEADPTAISGMRKRAHSRVEDFPCCDPTMDLMEAKDKEQA 542
QY 533 VFELMRTFKLPGEFKINDMVYDRDDENRPPVVPKAPKKKGKGRKRTTDEEAAPPEVD 592
Db 543 VFELCRRYEL-----QNPPEP-----LEIMGPE--- 565
QY 593 TAADVDADEGPKPEVPEERAMGGPNRYVWPLGCEEWLKPKEVKKVGGKVPVEEDE 651
Db 566 --YDQTRG-----YYPEGAEKRLTARKRRSRK-----EEVEED 597

RESULT 3
E83974
Hypothetical protein BH597 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: E83974
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Fujii, F.; H.
Nucleic Acids Res. 28: 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; MUID:20263314
A:Accession: E83974
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-322 <STO>
A:Cross-references: GB:AP001516; GB:BA000004; NID:gl0175192; PIDN:BAB06316.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2597

Query Match 10.8%; Score 377; DB 2; Length 322;
Best Local Similarity 29.1%; Pred. No. 2.4e-16;
Matches 105; Conservative 61; Mismatches 127; Indels 68; Gaps 10;

QY 207 RLYACLTNTYLRNAKPPIPSSRTCRMASIVDRHPLQFEDRPEHHLKNKPKSKPEQDEL 266
Db 4 QFGYVAMSMELANASP-----SKTMTAT-----QFEKIEDH-----EA 36
QY 267 GHKPVQELGLANARDIVKMLCWNEYKIGRFLRLSSEMFPFASHPV-HGYKLAFFASEVLA 325
Db 37 GLKRLERIAKTNLHNCRLKLLKUNLAQISFELSSKLPVNLHPLTEGKVELAIAEELQ 96
QY 326 EAGKVAELGHLRTHPCQFTQLGSPKPEVESAI RDLLEYHDELLSLKLPQGNRDVAM 385
Db 97 AVGEFASEHQMRIDFPHDFVVLNSEAKETIRSLQTLIDYHYKLLKLGMEIDRHR- - - -C 152
QY 386 LIHMGQFQDKAATLERKKNYARLSQSKNRLVLENDVGVTHDLPVCEELNIPMWL 445
Db 153 VLVHGGKRRKQVEACLGQFIENTASIPKSLMSLMLENDOKSYTIDVLYLGEKLAIPVVL 212
QY 446 DYHHNTCFDPAHLREGTLDISDPKLOERTANTYLRKGTGKQMHYSEPCDGAATPR- - - - 501
Db 213 DIHHVDLHRSKSLQE-----TWQRIATWEDSPILVKIHLSSPLSGEDDPRHNDY 263
QY 502 ---DRKRRPRVMTLPCCPD-----MDLMEAKDKEQAVFELMRTFKLPGEFKINDMVY 554
Db 264 INADR-----FIAPLHEIGADAVHLHVMLEAKKDLALFQLMK-----DLAEY 307
QY 555 D 555
Db 308 D 308

```

1 MPSSKRAALDTPQESSTFSTLDSAPSPARNLRSGNNILQPSSKDRO-----H 54
::
504 PSHRLKQA-----DSSSKSSSSSSSSSRSSRSDSSGSRHSPLRSQRDAQARTH 559
::
55 EKRSGEELAG-----RMCKDKANGICLR-----EKGQEGRGVKM 88
:: :

Db 560 ANPRGPRKMGSRSTSESRSRSRSASSNRKSLSPGVSRDSSSTSYTETKDPSSGQEV 619
 Qy 89 AIEGLARM-----ERRLQRTKQKQKQLEDGTPVPSVVSRRFTAYHHKSNARE- 141
 Db 620 ATPVPQVQVCEPKERTSTSSVQARLQSPESAERKHVTQRLQPERGSPKCEAEAE 679
 Qy 142 --AKEP---VLKTHSKD-----VEREABIGVD-----DVKMEPAATNIIEPD 180
 Db 680 PAATQPTSETQTHLPESERIHHTVEKEKEVMTDTSNRPNDV-----PEPMPADIQV 735
 Qy 181 AQDAEAERGAARPPAVNSSYLPLPWKGR-----GYACILNTYLRNAKP----- 222
 Db 736 SNDRPEGSVEDEKESLSPKSRKTSVVVSATKGVPAAGNSDPGPGPKRRWGA 795
 Qy 223 -----PIFSSCTCMASIVDRHRLQLEDEPEHLKKNPKSKPEODEL---GHKFFQEL 274
 Db 796 TQKKPSIITELSLKSLIPDIKLAGO---EAVVDLHADDSDRSEDETERNGDDGTHDK 852
 Qy 275 GLANARDIVKMLCWNEKYGIRFLRLSEMFPFASHPVHGYKLAPFASEVIAEAGRVAAEL 334
 Db 853 GLKICRTVTQVVPABQEQNGQREEEERKEPEAEPPVPQVSVVEVALPPPAEHEVKKVTL 912
 Qy 335 GRLUTH-----PGQFTQSGPRKEVVESEAIROLEYHDELL---SLKLKP 376
 Db 913 GDTLRRSISQOKSGVSTTIDDPVRTAQVSPPR-----GKTSNVHISNLVRPFTLGOLK 968
 Qy 377 EQONRDVAMIIHMGQFGDKAATLERFKN---YARLSQCKNRLVLENDVDGWTVDHL 432
 Db 969 ELLGRGTGLV-----EFAFWIDKIKSHCFVTYSTVEEAVATRTALHG--VKMPQSNP 1018
 Qy 433 LPVCEELNIPWVLYHHNICFDPFAHLREGTLDLSDPKLQRIANTWKRKGKIKOMHYSE 492
 Db 1019 KFLCADAEOQDELVDHRLGLVDRP-----SETKTEB-----QGIPRPLPPP 1060
 Qy 493 PCDGAVTPDRKRHRPVMTLPPCPDMDLMEAKDEAOAVE--LMRTFKLPQGFEXIND 550
 Db 1061 P-----PPVQPPHPRABOQEOERAVRQWABEREMERERTRS 1100
 Qy 551 MYPYDRD-----DENRAPPVAPKPKKGGKRRRTTDEAAEPEEVDTAADV-----KDA 601
 Db 1101 EREWDRKVREGPRSSRSRRRRKAKSEKKSEKKAQAEPPAKLLDDLFRTKAA 1160
 Qy 602 PRGPKVEPEERAMGPNRYVW-PL-----CCEWMLKPKREVKKGKVPVEVEDEG 652
 Db 1161 P-----CIYWLPLTDSIQVKEAERAKERKRKEQEEBEOKER 1202
 Qy 653 E 653
 Db 1203 E 1203
 RESULT 7
 T13564
 microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
 N:Alternate names: hypothetical protein BG:49E4.1
 C:Species: Drosophila melanogaster
 C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C:Accession: T13564
 R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
 submitted to the EMBL Data Library, April 1999
 A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
 A:Reference number: Z17689
 A:Accession: T13564
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-5327 <SPA>
 A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
 C:Genetics:
 A:Cross-references: FlyBase:FBgn0025392
 A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
 A>Note: BG:49E4.1
 C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 4.5%; Score 155.5; DB 2; Length 5327;
 Best Local Similarity 19.4%; Pred No. 0.56;
 Matches 151; Conservative 104; Mismatches 306; Indels 219; Gaps 34;
 Qy 3 SRKSKAAALDTQSSSTSF-----STLDSSAPSPARNLRRSGRNILQPSSE-KDRDHE 55
 Db 1378 TEKSPILASKETSRPESATGSKVEDTEQTKKKSPVPSR-----PESEADKKSP 1426
 Qy 56 KRSGE-----ELAGRMGKD---ANGH-----CLREGKEQE-----BGVK-- 87
 Db 1427 PASGEASRPESVAESVKDEAGKAERRESIAKTHKDESLDKAKQESRRRESLAESIAKE 1486
 Qy 88 -----NALIELAMERLRQATRKQKQLEEDCIPVPSVVSREPTAPYHH----- 132
 Db 1487 SGIDEKASALASKEARPESVTDKSKEPSRRESIAESLKAESTKDEKSAPPSKEASRPGSV 1546
 Qy 133 -KSTNAEREAKEP-----VLKTHSKDVP-REAE--TGVDVVVKMEPAATNIIEPDAQD 183
 Db 1547 VESVKDETEKSKESRRESIAESAKPIIEFREVSREPESVIDIKDESA-----KPSRRD 1601
 Qy 184 A--AERGAARPPAVNSSYLPLPWKGRGLGYACILNTYLRNAK--PPIFSSRTCRMASIV---- 236
 Db 1602 SPLASKEARPESV-----LESVKDEPIKSTESKRRESVAESFK 1640
 Qy 237 -----DHRHLPQED--EPEHHLKNKPD---KSKPQDELGHKFQVEL----- 274
 Db 1641 ADSTKDEKSPKTSKDISRPESAVENYMDAPFKETSRPESAVGSKDESKESRRESYK 1700
 Qy 275 -GLANARDIVK--MLCWNEKYGIRFLR-----LSSEMPPFAHPVHGYK 315
 Db 1701 DGAASRETSRPAESAKDAGADDLKELSRPESITQSKAAGSIKDEKSPASE--EASR 1758
 Qy 316 LAPFASEVIAEAGRVAAELGHLRTHPGQFTQSGPRKEVVESEAIROLEYHDELLSLKL 375
 Db 1759 PASVAESVKDEAKSKESRRESVAESKSPLSKEASRPAESIKD-----EAEKS 1810
 Qy 376 PCOONRDVAMIIHMGQFGDKAATLERFKNRYARLSQCKNRLVLENDVDGWTVDHLLPV 435
 Db 1811 KEESRRESV-----AEKSPLSKEASRPAESVAESIKDEAKSKESRRESVAESKSPLSKEASRPA 1856
 Qy 436 CEELNIPWVLYHHNICFDPFAHLREGTLDLSDPKLQRIANTWKRKGKIKOMHYSEPCD 495
 Db 1857 AKSPLP-----SKEASRPAESVAESIKDEAKSKESRRESVAESKSPLSKEASRPA 1909
 Qy 496 GAVTPRD-----RRKHRPVMTLPPCPP-----DMDLMEAKDKEQAVFELMRTFKLP 544
 Db 1910 VAESI KDEAKSKESRRESVAESKSPLSKEASRPAESVAESIKDEAKSKESR----- 1964
 Qy 545 FEKINDWVYDRDDENRPPPV-----KAPKKKGKRRRTTD-----EEAAEPEEVD 593
 Db 1965 -ESVAESKSPLSKEASRPAESVAESIKDEAKSKESRRESVAESKSPLSKEASRPAV-- 2021
 Qy 594 AADVDAEPGPKVEPEERAMGPNRYVWPLGCEWMLKPKRKREVKKGKVPVEVEDEGE 653
 Db 2022 -AESIKDEAKSKESRRESV-----AEKSPLSKEASRPAESVAESIKDEAE 2067
 RESULT 8
 S56231
 hypothetical protein YFL023w - yeast (Saccharomyces cerevisiae)
 C:Species: Saccharomyces cerevisiae
 C>Date: 02-Sep-1995 #sequence_revision 19-Oct-1995 #text_change 29-Oct-1999
 C:Accession: S56231
 R:Murakami, Y.; Naitou, M.; Hagihara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sas
 submitted to the EMBL Data Library, May 1995
 A:Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces
 A:Reference number: S56186
 A:Accession: S56231
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-796 <MUR>
 A:Cross-references: EMBL:D50617; NID:g836685; PIDN:BAA09215.1; PID:d1009856; PID:g836

C:Genetics:
A:Map position: 6L

Query Match 4.2%; Score 146; DB 2; Length 796;
Best Local Similarity 19.2%; Pred. No. 0.18;
Matches 108; Conservative 97; Mismatches 183; Indels 174; Gaps 28;

```

QY 4 RSKAAALDTPSESSST-FSTLDSSAPARNLRSGRNILOPSSEKBDHEKRSGEEL 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 326 KSPAENSNDKKKKQKGVFASSLEI---HEVENLKENKQKQMSFVPMVETQESTG--I 380
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 63 AGRMGKDANGHC-----LREKGEQEGVYKMAIEGLARMERLQATRKQKQKOLEEDGI 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 381 ANKWTSEDFDGLFAKMLGVQEADEVHEKYKEELINQERLEGASNSNRTR----- 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 117 PVPVSVSRF-----PTAPYHHKSTNAEREA--KEPVLTHTSKDVEREAEG--- 161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 433 -----VSRFRKDRASKKENTLSTFKQETTRSVENVEVKEPVV---GDIIEKEPVVGDVI 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 162 -----VDDVVKMEPAATNIIEPEDA-QDAARERG-----AARPPAVNS-S 198
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 485 EKEPVGDVIEKEPAVTDIVEREPAVNDIVERKPVVGDIIIEKPTINDIVEKEPEINSKS 544
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 199 YLPLPWKGRGLGYACLNLYLRNAKPPIFSSRTCRM---ASIVDRHRPLQPEDEPEHLKKN 255
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 545 EFETPPKKK-----KLASLOKP-RSSKSMKKKFKDPKILENISDDDDYDDDDGHNKLL 595
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 256 PDKSKEPQDELGHKF---VOELGLANARDIVKMLCWNEKYGIRFLRLSSEMFPFASHPVH 312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 596 SNKSKNNTDE-QDKFPSKIOEVSRSMAKT-----GATVGSPEVRITNVDY----HALG 643
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 313 G-----YKLAPFASEVLAEAGRAVAELGHLRTHPGQFTOLGSPRKEVVESAIRDLE 364
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 644 GNLDMDWKAYSGLYDDLEDPGTIVEKL-----EDFK 677
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 365 YHDELLSKLLPEQONRDVMIHMGQFGDKAATLEREKRYARLSQSCKNRLVLEND 424
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 678 EYNKQVELLR---DEIRDF-----QLENKPVME-----EEN 707
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 425 VGTWVHLLPVCEELNIPVLDYHHNICFDPAHLREGTLDISDPKLOERIANTRKKG 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 708 DGNVMDII---EHEFPESYTTNDEVALHPGLQAE-EVAIEYRRLKEATASKWQS--- 759
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 485 KQKHVSEPC---DGAVTPRDR 503
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 760 -----SSPAAHTEGELEPIDK 775
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Query Match 4.1%; Score 141.5; DB 2; Length 1110;
Best Local Similarity 17.6%; Pred. No. 0.53;
Matches 132; Conservative 102; Mismatches 275; Indels 239; Gaps 24;

QY 2 PSRKSAAALDTP-----QSESTFSSTLDSSAPARNLRSGRNILQPSSEKDRD 53
: :

```

DB 429 PSAKPAKAKVEKKVSKPKETIKVSEPIISAQLDITDLEDLAQQEVMKAAAPVWSAEKDEE 488
QY 54 HE-----KRSGEELAGRMCKKDANGHCLRECKQOEBOGVKMAIEGLARMERRL 100
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 489 EEEEEKEEEEEAAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 548
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 101 ORATRKQK-----KOLEEDGI-PVPVSVSRFPTAPYHHKSTNAEREA 142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 549 EAAEEEEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 608
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 143 KPVLTHTSKDVEREAEGVDDVVKMEPAATNIIEPEDAQAARGAARPPAVNSYSLPL 202
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 609 EEEVEAEATKE-EVEAAEAEVEB-----EGAAEEAEAEAEAEAEAEAEAE 651
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 203 PWKGRGLGYACLNLYLRNAKPPIFSSRTCRMASIVDRHRPLQPEDEPEHLKKNPKSKE 261
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 652 -----KAKTQEAEEVEDEEAAEAEAEAEAEAEAEAEAEAEAEAEAE 700
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 262 --PODELGHKFVOELGLANARDIVKMLCWNEKYGIRFLRLSSEMFPFASHPVHGYKLAPF 319
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 701 DAEEDEAEVEEVEKE-----EEVTK-----SD 721
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 320 ASEVLAEAGRAVAELGHLRTHPGQFTOLGSPRKEVVESAIRDLEYHDELLSKLLPBOQ 379
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 722 AEEAAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 769
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 380 NRDAVMIHMGQFGDKAATLEREKRYARLSQSCKNR-----LVLENDYGV 427
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 770 AKEA-----SDDEKPEEEVKESEAPVAEAKKAPKPAKPKKAPAKVES----- 814
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 428 TVHDLPLVCEELNIPM--VLDYHHNICFDPAHLREGTLDISDPKLOERIANTRKKG 485
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 815 -----PTSEPEDEPKAEVVE-----KKGAEAPKPAKPAKAAAKK-EAKPE 854
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 486 OKMHYSE-----PCDCAVTPRDRRRHRP---RVMTLPCCPDMDLMI 524
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 855 KEEPEESPTEEPKKPAKAPKAPKPAKPAKAEAEKPEPAKPAKPAKPAKPAKPAK 914
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 525 EAKDKEQAVFELMRTFKLPGEKINDMVPYD-----RQDENRPPAPVPAKPKKK 573
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 915 EKEDDEEEE-----EEVEVKPEIDAKPVKSKPAPAKAEEDPEPKAKQPKPKR 962
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 574 GKKRRTTDEFAAEP-----EEVDTAADDVKADEGPKPEVEPEERAMGPGYNRYW 625
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 963 KPAPKPEPEDEKAEPAKAEKHSVPEERKPKIEIAKPAKPAKADKEPEA----- 1011
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 626 LQCEEWLAKPKREVKKGVPEEVEDEGE 653
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1012 -----AEPKKIEVKVKVKVTKKVVEEIE 1033
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 10

T17292
hypothetical protein DKFZp434F192.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T17292
R:Anson, W.; Wirginer, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A:Reference number: Z18729
A:Accession: T17292
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-727 <ANS>
A:Cross-references: EMBL:AL117534
A:Experimental source: adult testis; clone DKFZp434F192
C:Genetics:
A:Note: DKFZp434F192.1

Query Match 4.0%; Score 140; DB 2; Length 727;
Best Local Similarity 19.4%; Pred. No. 0.38;
Matches 131; Conservative 92; Mismatches 213; Indels 238; Gaps 32;

```
QY 62 LAGRMCKDANGCHLREGKE--OBEQVKMAI--EGLARMERRLORATKROKKOLEEDCIP- 117
Db 21 LENLKGNKTGG-LFQKTPRLKRLQAOIIVPLKPDVNTYK--EAEKENIVEOSIPS 77
QY 118 -----VPSVSRFPTAPYHHKSTNAEAREAKEPEVLKTHSKDVERE--AEIGVDVV-- 166
Db 78 NACSSLEVAEAIISRKTPAQORSLRSAGKDLQEKKEHHVKKAKRCATPVVIIDILPS 137
QY 167 KMEPAANIIEPEDA-----ODAAERGAARPAVNSYLLPLPWKGRGLYACLTNYLRNKP 222
Db 138 KMKVSNKKKPEEGSAHODTEKNASSPEKA-----KGRHTVPCM-----P 180
QY 223 PTFSSRTRCMASIVDHRHPLQFDEPEHHLNKKPKDSKEPD--ELGKFKVOELGLANARD 281
Db 181 P-----AKQFLKSTEQELEKSMKKQOE----- 204
QY 282 IVKMLCWNEKYIGIFRLSLSSEMPFASHPVHGKYLAPFASEVLAAGRAAAELGHRLLTTH 341
Db 205 VVMRKKNEEF-----KKLALAG----- 222
QY 342 PGQFTQLGSPRKEVESAIRDLEYH---DELLSLLKLPEQONRDVMI IHMGQFGDKAA 398
Db 223 -----IGQPVKKSVSQVTKSFDFHFTDERIKOHPKNOEYKEV-----NFTS 265
QY 399 TLERFKNYARLSQSC-----KNRLVLENDVGVTHDHLPLVCGEELNIPMVLDY 447
Db 266 ELRKHPSPARVTKGCTIVKPFNLSQSKRTFDE-----TVSTYVPLAQO-----VEDF 314
QY 448 HHNHCIFDPAHLREGLTIDSPKLOERIANTRKKGIKOMHYSEPCDGAVTPRDRKKHR 507
Db 315 HKRTP--NRYHLRSKKDDIN-----LLPSKSSVTKICROPOTPVLOTKKR 357
QY 508 PRVMTLPPCPDMDLMEAKDKQAVFELMRTP-----KLPGFEKINDVVPYD 555
Db 358 ARVT---CASTAELEAELEKLOQYKFRARELDPRLEGGPLPKKPPVKPTEPIGED 414
QY 556 -----RDDE-----NRAP-----PVKAPKKGKGRKR 579
Db 415 LEIEKRIQERESKKTBDEHFEHSRCPCKTILEDVGVPEKKVLPTVTPKSPAFALKNR 474
QY 580 ---TTDEAAPEPEVDTAADVK--DAPEGPKEVPEERAMGGPNRVVWPLGCEBWLKP 634
Db 475 IRMPTKEDEDEPVTIKAPQVPHYGVPFKP-QIPEARTVEICPFS--FDSRDKRQLQOK 531
QY 635 KK--REYKKGKYPE 646
Db 532 EKKIKELQKGEVPK 545

RESULT 11
T33100
hypothetical protein H11E01.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
R:Geisel, C.; Wansley, P.
submitted to the EMBL Data Library, May 1998
A:Description: The sequence of C. elegans cosmid H11E01.
A:Reference number: Z21282
A:Accession: T33100
C:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1510 <GET>
A:Cross-references: EMBL:AF067222; PIDN:AACT17017.1; GSPDB:GN00028; CESP:H11E01.3
A:Experimental source: strain Bristol N2; clone H11E01
C:Genetics:
A:Gene: CESP:H11E01.3
A:Map position: X
A:Introns: 73/2; 120/1; 178/3; 227/2; 262/1; 300/2; 345/1; 1430/3; 1468/2

Query Match 4.0%; Score 139.5; DB 2: Length 1510;
```

```
Best Local Similarity 19.5%; Pred. No. 1-1;
Matches 151; Conservative 86; Mismatches 302; Indels 237; Gaps 38;
QY 8 AAALDTPQESSTFST-----LDSSAPSPARNLRRSGRNLIOPSSKED 51
Db 660 ASERDVPESAELVQSENFENFHAHVHSPHASEHIDEALPSPVQSVHSS----- 707
QY 52 RDHEKRS---GEELAGR---MMGKDANGCHLREGKEQEGVKMALEGLARMERLORATK 105
Db 708 HDHIDRRSPVASEPAARSPVQSSRTSEHFHRGEVPOSPSSNQFHSHSELEE-----A 761
QY 106 RQKKLEEDGIPVPSVSVSRF---PTAPYHHKSTNAEAREAKEPEVLKTH--SKDVEREAEIG 161
Db 762 RQSPVTNQESVHSPHASEHFEHREWVP--HSPAASQEEHGRSPSVHSPHSEHFEHHEEAQ 820
QY 162 VDDVVKMEFAA-----TNIIPEDAQD---AERGAARPPAVNSYLLPLPWKGR 207
Db 821 HSPVASQEEAARSPSVHSHASEHFEHHEEAQDPSVAQEAARSPSVHSHASEHSEHR 880
QY 208 LGY-----ACLTNLYRNAKPPDIFSRTCRMASIVDHRHPLQFDEP-----EHLKLNK 255
Db 881 QEIQHSAPAASQNEAARS--PSVHSHASE--HIENHGESLQ---SPVASMAGSEHH--NM 931
QY 256 PDKSKEPQDELCHKFVQBELGLANARDIVKMLCWNEKYIGIFRLSLSSEMF-----PFAS 308
Db 932 AESSEYTTTSE-----KEISPSI-----FSSHTSEQEFOQOQSONSPVAS 968
QY 309 HPVHGKYLAPFASEVLAAGRAVAAFLGHRLLTHPCOFTQLGSPRKEVVE-----AIRDLE 364
Db 969 E--RDNRSPTFESSVTMOA---AAPLSPAASDHAEQARE---SPSFEAPSLHLSOLSNLE 1021
QY 365 YHDELLSLLKLPEQONRDVMI IHMGQFGDKAATLERFKNYARLSQSKKNRLVLEND 424
Db 1022 HDDENSAAVVEAGQEPATQSPIPLEQGRF--ERAASV-----NSYOASES-----FENQE 1069
QY 425 VGWTVHDL---LPVCEELNIPMVLDY-----HHNHCIFDPAHLR 460
Db 1070 ATVVEHHAGALQSPVASEKEVSEAI SQPAEPAAPSPISSEVEAHSEHHHHPQEEPIVQT 1129
QY 461 EGTLDLSD--PKLQERIANTRKKGIKOMHYSEPCDGAVTPRDRKKHRPRVMTLPPCPPD 519
Db 1130 SHHIDLTVDPRSPAFSMN-----DSHHEDQLSHEAAP-----HSPVAFTEPRVP-- 1173
QY 520 MDLMEAKLQKQAVFELMR-----TFKLPGEKINDVVPYDRDRNRPAAPPVKAP----- 569
Db 1174 ---SVHTSCTEESEFEVQHAGLTQEYAAKSPTEMDLYNPSRDNQESPVNSETAIPIV 1230
QY 570 -----KKKKGKRRKRTTDEAAPEEVDTAADDV-----KDAPE 603
Db 1231 DRIDLEAVHNSPTKESTSPVTNVNVEVRSADAETHLASHQNFEDPSPHLVDESEVPN 1290
QY 604 GPKEVPEEER---AMGGPNRVYWP---LGCEEWLKKPKKREVKKGKVPVEEDEGE 653
Db 1291 NFETVPEQEAAPSODAMMTSIYQPGSDTETREW-----DEGE 1328

RESULT 12
C81169
IgA-specific metalloendopeptidase (EC 3.4.24.13) NM00700 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
A:Variety: strain MC58 serogroup B; strain 811139
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: C81169; S61326
R:Teitelman, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.; Qi, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755
A:Accession: C81169
A:Molecule type: DNA
A:Residues: 1-1815 <TET>
```

A: Cross-references: GB:AE002424; GB:AE002098; NID:g7225923; PIDN:AAF41117.1; PID:g7225923
A: Experimental source: serogroup B, strain MC58
R: Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A: Title: Comparative characterization of the iga gene encoding IgA1 protease in *Neisseria meningitidis* serogroup B strains
A: Reference number: S61314; MUID:95302961
A: Accession: S61326
A: Status: preliminary; nucleic acid sequence not shown
A: Molecule type: DNA
A: Residues: 53-548 <LOM>
A: Cross-references: EMBL:X82477; NID:g732856; PIDN:CAA57860.1; PID:g732857
C: Genetics:
A: Gene: NMB0700
C: Superfamily: IGA-specific metalloendopeptidase
C: Keywords: hydrolase; metalloproteinase

Query Match	4.08;	Score 139;	DB 2;	Length 1815;
Best Local Similarity	18.78;	Pred. No. 1.4;		
Matches 125; Conservative	94;	Mismatches 268;	Indels 180;	Gaps 26;

Qy	38	RSGRNILPQSEKDRDHEKRSGEELAGRMGKDANGH----	CLREGKROEEGVKMAIEGL	93
		: : : :	: : : :	
Db	916	QEGLNLFDASSVQDR-----	SRLSVSLANNHVLDLGALRYTIKTENGITFLNPY	964
		: : : :	: : : :	
Qy	94	ARMERLRQRA-----TKRKKOLEDG--IPVPVVVSREPTAPYHHKSTNAEREAKE	144	
		: : : : : : : :	: : : : : : : :	
Db	965	AENRRVKPAESPATINTASQAQAKTQTDGAQIAKPNTIVVAPPSPQANOAEAKRQOAKA	1024	
		: : : : : : : :	: : : : : : : :	
Qy	145	PVLTHSKDVIREA-----ELGVDDVVMKEPAANTNI I-EPEDAODAAERG	188	
		: : : : : : : :	: : : : : : : :	
Db	1025	EQVRQQAERKSALAKQKAERARELARTQAKOEGRSSAELARRHEKEREAELS	1084	
		: : : : : : : :	: : : : : : : :	
Qy	189	AARPPAVNVSYLPWPWKGRLGYACINTLYLRNNKPPIFSSRTCRMASIVDHRHPLO-----	243	
		: : : : : : : :	: : : : : : : :	
Db	1085	AKQVEAREAQALAVR-----RKAE----AEEAKROAELARRHEKEREAEE	1128	
		: : : : : : : :	: : : : : : : :	
Qy	244	-----FEDPEHLKNPKDSKE---PODEL-----GHKFVOE--LGLANARD	281	
		: : : : : : : :	: : : : : : : :	
Db	1129	LSAKORVGEERRQTAAQOPQRKRRAAPODYMAASQDPKPRGHRSVQONNVEIAQAQA	1188	
		: : : : : : : :	: : : : : : : :	
Qy	282	IVKMICWNKEYGIRPL----RLSSEMFPFASHPVHGYYKLAPFASEVLAEAGRVAELGHLR	338	
		: : : : : : : :	: : : : : : : :	
Db	1189	ELARRQQEERKAAELLAKQRAEAER-----EAQALAARKBAEAEAKRQAEALAHQ	1240	
		: : : : : : : :	: : : : : : : :	
Qy	339	TTHPGQFTQLGSPRKEVYES--ATRDLHEYDELLSLKLPPQNRDAMVIIHMGGQFGD	395	
		: : : : : : : :	: : : : : : : :	
Db	1241	EAE-RKAELSANQKAAEAQAALARQOK-----ALARQOEARKAAELAVKQKAETER	1293	
		: : : : : : : :	: : : : : : : :	
Qy	396	KAA TL--BRFKRNYARLSQSCKNRLVLENDVDVGWTVHDLLPVCEELINPMVLDYHHHNIC	453	
		: : : : : : : :	: : : : : : : :	
Db	1294	KTAELAKQRAAAAANKRQEARQTAE L-----	1320	
		: : : : : : : :	: : : : : : : :	
Qy	454	FDPAHLREGTLDISDPKLOER-----IANTWKRGKI Q-----KMHYSEPCD	495	
		: : : : : : : :	: : : : : : : :	
Db	1321	---ARRQGAERQAALSQAQKAETDREAESA KRAAEHEEHQAQAOSQPQRKRRAAFQD	1377	
		: : : : : : : :	: : : : : : : :	
Qy	496	GAVTPRDRKRHRPVMTLP-PCPPDMDLMI EAKDKQAVFE LMRTFKLP G-FEKINDMPV	553	
		: : : : : : : :	: : : : : : : :	
Db	1378	YMAASQNRPKRGRGSTLPAPSPSFD-----SSAYAAPRALHNPDMWYENDYEELP	1428	
		: : : : : : : :	: : : : : : : :	
Qy	554	YDRDDENRAPPVY---KAPKKKCGKRRTTDEE-----AAPEVDTAADVDKADPEG	604	
		: : : : : : : :	: : : : ~ : : : :	
Db	1429	LDALEDNMVSESVDTS DKOPODNT ELHKKYENDYBEI PLDALEDDESVDTS DKQPOD	1488	
		: : : : : : : :	: : : : : : : :	
Qy	605	PKEVPEE	611	
		: : : :	: : : :	
Db	1489	NTELHEK	1495	
		: : : :	: : : :	

RESULT 13
 T28657
 blacklack protein, microtubule associated - American bird grasshopper

RESULT 13
T28657
blackjack

Db 1326 TSLATQKOSSTTGKVP 1342

RESULT 14

T42727

proliferation potential-related protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 02-Sep-2000

C:Accession: T42727

R:Witte, M.M.; Scott, R.E.

submitted to the EMBL Data Library, November 1998

A:Reference number: 222246

A:Accession: T42727

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1560 <WIT>

A:Cross-references: EMBL:U03913; NID:g3858884; PID:g3858885; PIDN:AAC72432.1

A:Experimental source: strain Balb/C

C:Genetics:

A:Gene: P2P-R

C:Function:

A:Description: involved in hnRNP association and Rb1 binding

C:Superfamily: RING finger homology

F;57-107/Domain: RING finger homology <RRN>

Query Match 3.9%; Score 137; DB 2; Length 1560;
Best Local Similarity 19.3%; Pred. No. 1.6;
Matches 157; Conservative 100; Mismatches 326; Indels 232; Gaps 33;

Qy 2 PSRKSAAALDTP-QSESSTFSSTLSSAPARNLRSGRNILQPSSEKDRDHEKRSGE 60

Db 749 PSREDATVDRDEMDAESITFKVSVD-----KDKREKDKPKVSKDKTKRSKSDGS 797

Qy 61 ELAGRMGMKDANGHCLREGQEQEGVGMATGELARMERLRQATKOKKOLEBDGIPVPS 120

Db 798 ATAKK-----DNVLPKSGQKQGVDRKSPRSEPPPLKKA-----KEEATKIDS 842

Qy 121 V-----VSRFTAPYHHKSTNAEREA-----EPVLKTHSKDVEREAIEGVDDV 165

Db 843 VKPSSSOKDEKVTGTGRKA--HKSAAKEHQEAQPAKDEKVKKDKCSKDKIASEKPSKDEK 900

Qy 166 VKMPEAATNII-----EPEDAQAAERGAARPPAVNSYLLPLPWKGLGYACLN 214

Db 901 AK-KPERNKLLDSKGEKRRKTEESVDKDFESSMKISKVEGTEIVKSPKMGEDVE 959

Qy 215 TYLRNAKPPITFSSTCRMASIVDHRHPLOFEDPEPHHLKNKPKDS--KPPQDEL---GHK 269

Db 960 KLRTPEKDKIASSTTPPAKKIKLNR-----ETGKKIGNAENASTTKPESEKLESTSSK 1012

Qy 270 FVQELGLANA-----EPEDAQAAERGAARPPAVNSYLLPLPWKGLGYACLN 214

Db 1013 IKQEVKVKAKRVKAGSESSSTLVDTYSTSTGSGPVKSEKTDITRIVIKM--EE 1069

Qy 292 YGIRFLSSEMPPFASHVHGKYLAPFAGE-----VLADAGRVAAELGHLRTHP- 342

Db 1070 YNDNTAPAEVDVLIHIVQPSKDKDDFESEEDVKTQPIQSVGKPSSTII-KNVTTKPS 1128

Qy 343 --GOFTOLGSPRKEVVB-----SAIRLDYHDELLSLKLPQEQNRDAVMTIIMG----- 390

Db 1129 ATAKYTEKESEQEKLOKLPKASHELMQH-ELRSKSGSASSKGRADREHSGSKDNP 1187

Qy 391 -----GQFGDKAATLERFKR--NYARLSQSKNRLVLENDVV--GWTVHDLPLVPCBELNI 441

Db 1188 DKRKSQAQPKESTVDRLSPQGHFKTLSSQSKETRTSEKHESVRGSSNKDETTPGRDK--- 1244

Qy 442 PMVLDYHHNICDPHAHLRGGTL-----DISDPKLOERTANTWKKIGIKQKMH 489

Db 1245 --KVYDSRYSKSKRDERGELARKDPSPPKESLSGOKSKLREERDLPKKGAESRKS 1302

Qy 490 YSEPCDGAVTTPDRRKH-----RPRVMTLP----- 514

Db 1303 NSSP-----PROKKPHDKAPYETRRKPCETKPVDKISKREKHAEEARNKDSGGK 1356

Qy 515 -PC---PPLMDLMIKAKUKEQAVFELMRTEKLP-----GFEKINDMVPYDRDD 558
Db 1357 LFCILNPPULPW-----EKELAVGQVKSVAKPQLSHSSRLSSDLTRETDEAAFEPTY 1411
Qy 559 ENRPAPPVKAPKKK-----GCKRRRTTDFEAAAPFEVDTAAD---DVKDAPEGKPEYPE 610
Db 1412 NESDSSESNVSVKEEAVASISKDLKEKTEKAKESLTVATASQPCADRSQSQSSIVSVSPS 1471
Qy 611 EERAMGGPYNRVY-----WPLGCEEWLKKPKREVKK 641
Db 1472 RSHSPSGQSTRSHSSASSAGSQDKSKKKKKKKKK 1506

RESULT 15

R48666

cell proliferation antigen Ki-67, short form - human

C:Species: Homo sapiens (man)

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 07-May-1999

C:Accession: B48666

R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Ge

A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, u

ins.

A:Reference number: A48666; MUID:94043435

A:Accession: B48666

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2897 <SCH>

A:Cross-references: EMBL:X65551

C:Superfamily: kinase interaction domain homology

C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat

F;29-91/Domain: kinase interaction domain homology <KIH>

Query Match 3.9%; Score 137; DB 2; Length 2897;
Best Local Similarity 19.9%; Pred. No. 3.6;
Matches 146; Conservative 92; Mismatches 296; Indels 200; Gaps 30;

Qy 5 KSKAALD-----TP--QSESSTFSSTLSSAPSPA-----RNLRSGRNILQPS 48

Db 2069 KEKALELDVGFELFQTPGHTESMTDDKITEVCSKSPQESFKTSRSKQRLKPLV 2128

Qy 49 EKDRDHEKRSGEELAGRMGMKDANGHCLREG-----KEQESCVKMAIEGLARME-RLQR 102

Db 2129 KVDMEELAVSKLT-RTSGETTQTHTEPTGDSKSIKAFESKPOILDPAASVTSRRQL 2187

Qy 103 ATKROKKOLEEDGIPVPSVSRFTAPYHHKSTNAEERAEKPV----- 146

Db 2188 KTRKEKARALEDLVDVKELFS----APGHTESMTIDKNTKIPCKSPPELTDITATSKR 2243

Qy 147 -----LKTHSKDVERAEI-GVDDVVKMEPA-----TNII-----EPEDAQAAE 186

Db 2244 CKTPRPRKEVKEELSAVERLTQTSQSTHTHKEPASGDGKVLKQRAKKKPNPVEEPS 2303

Qy 187 RGAARPPAVNSYLLPLPWKGLGYACLNLYLRNAKPPILFSSRTCRMASIVDHRHPLOFED 246

Db 2304 RRRPRAPKEAQ---PLEDLAGTELSETSGHTQESLTACKATK---IPCESPLEWVD 2356

Qy 247 ---EPEHILKNKPKSKPEQDELGHKFWOELGLANARDIVKMLCWNEKYGIRFLRSLSEM 303

Db 2457 TJATSKRHLTRVQVQVKEEPSAVKFTQTSQ---ETTDADKEPAGEDK-GIKALKESAKQ 2413

Qy 304 PPFASHPVILGKLPFAFSEVLAE-----AGRVAAELGH-----RLTHP-----GOFT 346

Db 2414 TPAPAASTVGSRRRPRAPRESAQALDLAGKDPDPAAGHTESMTDDKTKIPCKSSPELE 2473

Qy 347 QLGSPRKEVSEAI-RDLEYHDELISLKL-----PEQON 380

Db 2474 DTATSSKRRPRTRAKQVEKVELLAVGLKLTQTSGETTHDKEPVGEGKTKAFKQPAKRN 2533

Qy 381 RDVMTIIMGQ---FGDKAATLERFRFNARLSQSCKNRLVLENDIVGTVTHLLPVCE 437

Db 2534 VDAEDVIGSRQPRAPKEKAQPLEDL-ASFQELSQT----- 2568
Qy 438 ELNIPWVLDYHHHNICFDPAPHLRE-----GTLDISDP-KLOERIANTWKRR 482
Db 2569 -----PGHTEELANGAADSFTSAPKOTPDGKPLKISRRLRAPK-- 2608
Qy 483 GIKQMHYSEPCDGAVTPRDRRKHPRVMT-LPPCPPDMDLMEAKDKQAVFELMRTEK 541
Db 2609 -----VEPVGDVSTRDPVKSQSKSNTSLPPLP-----FKRGGKDGSVTGTKRLRC 2655
Qy 542 LPGFEKINDMVPYDRDDENRAPPPVAPKAPKKKGKRRKTTDEEAAPPEEVDTAADDVKDA 601
Db 2656 MPAPPEIVEELPASKQORVAPRARGKSSEPVWIMKRSRLRTSAKRIBPAE-ELNSNDMKTN 2714
Qy 602 PEGPK-----EVPPEE 612
Db 2715 KEEHKLQDSVPENK 2728

Search completed: January 15, 2002, 13:52:31
Job time: 400 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2002, 13:49:55 ; Search time 120.01 Seconds
(without alignments)
369.718 Million cell updates/sec

Title: US-09-724-296-2_COPY_230_828

Perfect score: 3112

Sequence: 1 MLRLKRNQIKSRIVFTIL.....KRLTARKRSRKEVEDEK 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A_Geneseq_1101.*
- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
- 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
- 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
- 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
- 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
- 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
- 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
- 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
- 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
- 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
- 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
- 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
- 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
- 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
- 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
- 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
- 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
- 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
- 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
- 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
- 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
- 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3112	100.0	828	21 AAY44498	GST signal peptide
2	1979	63.6	371	21 AAY44499	S. pombe delta228-
3	1979	63.6	600	21 AAY44500	GST signal peptide
4	1011.5	32.5	656	17 AAW01618	Neurospora crassa
5	961.5	30.9	626	21 AAY44502	N. crassa delta228
6	961.5	30.9	626	21 AAY44503	B. subtilis delta2
7	230.5	7.4	294	21 AAY44505	D. radiodurans del
8	147.5	4.7	782	12 AAR06991	Polypeptide antige
9	140.5	4.5	612	22 AAB95546	Human protein sequ
10	140.5	4.5	733	19 AAW60785	Deletion mutant de
11	140.5	4.5	2441	21 AAB18161	Plasmodium falcipa

12	136.5	4.4	1639	19 AAW54145	P. falciparum synt
13	135	4.3	1341	21 AAY85657	Human Acinus L pro
14	134.5	4.3	753	19 AAW60788	Deletion mutant de
15	134.5	4.3	753	19 AAW60787	Deletion mutant de
16	134.5	4.3	1654	6 AAP50777	Sequence of the P1
17	134	4.3	2688	22 AAM40883	Human polypeptide
18	133.5	4.3	1588	15 AAR46608	Plasmodium falcipa
19	133.5	4.3	1663	15 AAR46608	Plasmodium falcipa
20	133	4.3	677	20 AAW85608	Secreted protein c
21	133	4.3	677	20 AAW85608	Human myosin heavy
22	132.5	4.3	2663	22 AAG64560	Human polypeptide
23	132	4.2	414	17 AAM39097	Saccharomyces cere
24	132	4.2	414	17 AAM39097	Amino acid sequenc
25	131.5	4.2	922	22 AAB30809	Human protein sequ
26	130.5	4.2	1197	22 AAB93218	Human protein sequ
27	130.5	4.2	1540	20 AAY07734	Human htlp2 protei
28	130.5	4.2	1674	19 AAW81169	Human BAZ1-alpha p
29	130	4.2	377	22 AAM18201	Peptide #4635 enco
30	130	4.2	377	22 AAM30696	Peptide #4733 enco
31	130	4.2	377	22 AAM05818	Peptide #4500 enco
32	130	4.2	615	21 AAG18697	Arabidopsis thalia
33	130	4.2	615	21 AAG50810	Arabidopsis thalia
34	130	4.2	633	21 AAG18696	Arabidopsis thalia
35	130	4.2	633	21 AAG50809	Arabidopsis thalia
36	130	4.2	662	21 AAG18695	Arabidopsis thalia
37	130	4.2	662	21 AAG50808	Arabidopsis thalia
38	130	4.2	1135	21 AAY84460	Amino acid sequenc
39	130	4.2	1257	22 AAB66926	RBP1 protein. Hom
40	129.5	4.2	431	20 AAY49137	Amino acid sequenc
41	129.5	4.2	489	22 AAB92515	Human protein sequ
42	129	4.1	1400	22 AAB83348	AAP-2 protein sequ
43	128.5	4.1	769	19 AAW60795	Deletion mutant de
44	128.5	4.1	769	19 AAW60796	Deletion mutant de
45	127.5	4.1	769	19 AAW60793	Deletion mutant de

ALIGNMENTS

RESULT 1
AAY44498
ID AAY44498 standard; Protein: 828 AA.

AC AAY44498:

DT 27-MAR-2000 (first entry)

XX GST signal peptide and S. pombe UVDE fusion protein.

DE GST signal peptide; ultraviolet damage endonuclease; UVDE; Uvelp;

KW UV irradiation; DNA damage; UV radiation damage; fusion protein;

KW skin cancer; glutathione-S-transferase.

XX Schizosaccharomyces pombe.

OS Synthetic.

XX Location/Qualifiers

FT Key Peptide

FT /label= GST_signal_peptide

FT Protein 230..828

FT /note= "S. pombe UVDE"

FT Misc-difference 11

FT /note= "Encoded by aaaa"

FT Misc-difference 85

FT /note= "Encoded by gtt"

FT Misc-difference 147

FT /note= "Encoded by gtt"

XX WO99063828-A1

PN 16-DEC-1999.

XX 08-JUN-1999;

XX 99WO-0512910.

PF

XX 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX (UYEM-) UNIV EMORY.
 PA Doetsch PW, Kaur B, Avery AM;
 PI
 XX
 XX
 DR WPI; 2000-116417/10.
 DR N-PSDB; AA229858.
 XX
 PT A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers
 XX
 XX Claim 7; Page 51; 133pp; English.
 XX
 PS The present sequence is a fusion protein comprising the GST signal
 CC peptide and full length ultraviolet damage endonuclease (UVDE) from
 CC Schizosaccharomyces pombe. UVDE genes are isolated by introducing a
 CC foreign cDNA library into a repair-deficient E. coli strain and
 CC selecting for complemented cells by UV irradiation of the transformants.
 CC This provides stable endonuclease fragments for cleaving a double-
 CC stranded DNA molecule that has a distorted structure resulting from UV
 CC radiation damage, a photoproduct, an abasic site, mismatched nucleotide
 CC pairing, platinum diadduct, an intercalated molecule or alkylation of a
 CC nucleotide. Uvelp can be used in compositions for internal or topical
 CC application and as a therapeutic agent for skin cancers.
 XX
 XX Sequence 828 AA;

Query Match 100.0%; Score 3112; DB 21; Length 828;
 Best Local Similarity 100.0%; Pred. No. 1.5e-245;
 Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRLKRNLIQSKRIVFTILKQAFKGNHPCVPSVCHTYSRHFCHLDPDTLKSLLPMSKST 60
 DB 230 mlrlkrlnlqskrlvftllkqafkgnhpcvpsvctitysrhfcpldtklslpmskst 289
 QY 61 TLSMLPQVNICANSFSAETPVDLKKNETELANISGPHKSTSTSTRKARSSKKKATDS 120
 DB 290 tlsmlpqvnicansfsaetpvdllkenetelanisgphkfstststrkarsskkalds 349
 QY 121 VSDKIDESVSYDSTHLLRRSRKPKVYNSSSESESEIQSKATKKVKEEVEEYVEE 180
 DB 350 vsdkidesvasydsthlrrsrkpkvynssseseseeqiskatkkvkeeeeyvee 409
 QY 181 VDEKSLKNSSDEEPVVPQLETPISKRRSRSSAKNLEKESTMLDDHAPREMFDCI 240
 DB 410 vdekslknssdeefvvpvpeqlptiskrrsrssaknlekesetmlddhapremfdcl 469
 QY 241 DKPIPWGRGLGYACILNTILRSKMERVFCSTCRITTTIORGLSVKOLGTONVLDLILKV 300
 DB 470 dkpwpgrglgyacilntilrsmkervfcsrtcrittiqrdglesvkglgtonvldliklv 529
 QY 301 EWNHNFTHMRVSSDLPPFASHAKYGYTLEFAOSHLEEVGKLANKYNHRLTMHPGQYQ 360
 DB 530 ewnhnfhgthmrsvssdlppfashakgytylefaoshleevgklankynhrltmhpggytq 589
 QY 361 IASPREVVVDSAIRDLAVHDEILSRMLNEQLNKDAVLIITHLGCTFEGKKTDLRFRKNY 420
 DB 590 iasprevvvdsairdlavhdeilsrmlneqlnkdavliihlgctfegkktldrfrkny 649
 QY 421 QRLSDSVKARLVLENDVSVSVDLLPLCOELNIPVLVDWHHHNINVPCTLRGSLDLMLPL 480
 DB 650 qrlsdsvkarlvlendvsvsvqdlplcqelnipvlvdwhhnnivpgtlregslldmpl 709
 QY 481 IPTRETWTRKGIQKHYSSESADPTAISGMKRAHSDRVDFPPCDPTMDLMEAKEKE 540
 DB 710 iptretwtrkgitqkhyssesadptaisgmkrrahsdrvdfppcdptmdlmeakeke 769
 QY 541 QAVFELCRRYELQNPCCPLEITMGPEYDQTRDGYYPGCAEKKRLTARKRSRKEVEEDEK 599

Db 770 qavfelcrryelnppcpleimgpeydyqrdgyypgcaekrltarrrsrkeveedek 828

RESUIT 2
 AAY44499
 ID AAY44499 standard; Protein; 371 AA.
 XX
 AC AAY44499;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE S. pombe delta228-UV damage endonuclease.
 XX
 KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
 KW glutathione-S-transferase signal peptide; uvel+ gene product;
 KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
 KW abasic site; platinum diadduct; mismatched nucleotide pairing;
 KW nucleotide alkylation; skin cancer.
 XX
 OS Schizosaccharomyces pombe.
 XX
 PN WO9963828-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX
 PA (UYEM-) UNIV EMORY.
 XX
 XX Doetsch PW, Kaur B, Avery AM;
 DR WPI; 2000-116417/10.
 DR N-PSDB; AA229859.
 XX
 PT A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers
 XX
 PS Claim 13; Page 53; 133pp; English.
 XX
 CC The present sequence is delta228-UV damage endonuclease. Delta228-UVDE is
 CC a 288 amino acid deletion of the N-terminal of the S. pombe uvel+ gene
 CC product. This is expressed in frame with a GST leader sequence to
 CC generate a fusion protein. This provides stable endonuclease fragments
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, platinum diadduct, an intercalated
 CC molecule or alkylation of a nucleotide. Uvelp can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 SQ Sequence 371 AA;

Query Match 63.6%; Score 1979; DB 21; Length 371;
 Best Local Similarity 100.0%; Pred. No. 1.9e-153;
 Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 DDHAPREMFDCIADKPIPWGRGLGYACILNTILRSKMERVFCSTCRITTTIORGLSVKOL 288
 DB 1 ddhapremfdcdkpipwgrglgyacilntilrsmkervfcsrtcrittiqrdglesvkgql 60
 QY 289 GTQNVLDLILKLVENHNFTHMRVSSDLPPFASHAKYGYTLEFAOSHLEEVGKLANKYN 348
 DB 61 gtqnvldlilklvenhnhfhgthmrsvssdlppfashakgytletaqshleevgklankyn 120
 QY 349 HRLTMHPGQYQYIASPREVVVDSAIRDLAYHDEILSRMLNEQLNKDAVLIITHLGCTFEG 408
 DB 121 hrltmhpggytqiasprevvvdsairdlayhdeilsrmlneqlnkdavliihlgctfeg 180
 QY 409 KKETLDRFRKNYQRLSDSVKARLVLENDVSVSVDLLPLCOELNIPVLVDWHHHNINVP 468

Db 181 kketldfrknygrlsdsvkariwlvnddsvsqdlplcqlnplvldwhhniwpg 240
 Qy 469 TLREGSLDLMLPIPTIRETWTRKGIQKOHYSESADPTAISGMKRAHSDRVDFPPCDP 528
 Db 241 tlregslldmpliptiretwtrkgitqkghysesadptaisgmkrhahsdvdfppcdp 300
 Qy 529 TMDLMIEAKEKEQAVFELCRRYELQNPCCPLEIMGPEYDQTRDGYPPGAEKRLTARKRR 588
 Db 301 tmdlmieakekedavfelcrryelnppcpelmgydqdtrdgyppgaekrltarkrr 360
 Qy 589 SRKEEVEEDEK 599
 Db 361 srkeeeveedek 371
 RESULT 3
 AAY44500
 ID AAY44500 standard; Protein: 600 AA.
 XX
 AC AAY44500;
 XX
 DT 27-MAR-2000 (first entry)
 DE GST signal peptide and delta228 S. pombe UVDE fusion protein.
 XX
 KW GST signal peptide; glutathione-S-transferase signal peptide;
 KW ultraviolet damage endonuclease; UVDE; delta228-UVDE;
 KW repair-deficient E. coli strain; UV irradiation; DNA damage;
 KW UV radiation damage; photoproduct; abasic site; platinum diaduct;
 KW mismatched nucleotide pairing; nucleotide alkylation;
 KW fusion protein; skin cancer.
 XX
 OS Schizosaccharomyces pombe.
 XX
 FH Synthetic.
 XX
 Key Location/Qualifiers
 FT Misc-difference 11 /note= "Encoded by aaaa"
 FT Misc-difference 85 /note= "Encoded by gtt"
 FT Misc-difference 147 /note= "Encoded by gtt"
 FT
 XX
 PN W09963828-A1..
 XX
 PD 16-DEC-1999.
 XX
 XX 08-JUN-1999; 99WO-US12910.
 XX
 XX 08-JUN-1998; 98US-0088521.
 XX 18-MAY-1999; 99US-0134752.
 XX (UYEM-) UNIV EMORY.
 XX
 XX Doetsch PW, Kaur B, Avery AM;
 XX
 XX WPI: 2000-116417/10.
 XX N-PSDB: AAZ29860.
 XX
 XX A new truncated ultraviolet damage endonuclease for treatment of skin
 cancers -
 XX
 XX Disclosure; Page 56; 133pp; English.
 XX
 CC The present sequence is a fusion protein comprising the GST signal
 CC peptide and delta228 truncated ultraviolet damage endonuclease (UVDE)
 CC from S. pombe. Delta228-UVDE is a 288 amino acid deletion of the
 CC N-terminal of the S. pombe uvrl+ gene product. This is expressed in frame
 CC with a GST leader sequence. Stable endonuclease fragments can be produced
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, platinum diaduct, an intercalated

CC molecule or alkylation of a nucleotide. Uvrlp can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 SQ Sequence 600 AA;
 Query Match 63.6%; Score 1979; DB 21; Length 600;
 Best Local Similarity 100.0%; Pred. No. 3.8e-153;
 Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 229 DDHAPREMPDCLDKPIPWGRGLGYACLNITILRSMKERVFCSTCRITTTIQRDGLSEVKOL 288
 Db 230 ddhapremfddcldkplpwgrgrlgyaclnitilrsmkervfcsrtcritttiqrdglesvkql 289
 Qy 289 GTQNVLDLILKLVENHNFCIHFMVRVSSDLFPFASHAKYGYTLEFAQSHLEEVCKLANKYN 348
 Db 290 gtqnvldlilkvewnhnfcghfmrvssdlfpfashakgytylefaqshleevgkllankyn 349
 Qy 349 HRLTMHPGQYTOIASPREVVVDSAIRDLAYHDEILSRMKLNQKNDAVLIIHLGGTFEG 408
 Db 350 hrltmhpgqygtiasprevvvdsairdlayhdeilsrmlkneqlnkdavliihlgttfeq 409
 Qy 409 KKETLDRFRKNYQRLSDSVKARLVLENDVSVSQDLPLCQELNIPVLVDWHHNIWPG 468
 Db 410 kketldfrknygrlsdsvkarlvlenddsvsqdlplcqlnplvldwhhniwpg 469
 Qy 469 TLREGSLDLMLPIPTIRETWTRKGIQKOHYSESADPTAISGMKRAHSDRVDFPPCDP 528
 Db 470 tlregslldmpliptiretwtrkgitqkghysesadptaisgmkrhahsdvdfppcdp 529
 Qy 529 TMDLMIEAKEKEQAVFELCRRYELQNPCCPLEIMGPEYDQTRDGYPPGAEKRLTARKRR 588
 Db 530 tmdlmieakekedavfelcrryelnppcpelmgydqdtrdgyppgaekrltarkrr 589
 Qy 589 SRKEEVEEDEK 599
 Db 590 srkeeeveedek 600
 RESULT 4
 AAW01618
 ID AAW01618 standard; Protein: 656 AA.
 XX
 AC AAW01618;
 XX
 DT 20-AUG-1997 (first entry)
 DE Neurospora crassa DNA repair enzyme.
 XX
 KW DNA repair enzyme; cyclobutane-type dimer; (6-4)-bound product;
 KW exposure; UV light; ultraviolet; suntan cream; prevention; treatment;
 KW skin cancer.
 XX
 OS Neurospora crassa.
 XX
 PN JP08266275-A.
 XX
 PD 15-OCT-1996.
 XX
 XX 29-MAR-1995; 95JP-0094137.
 XX 29-MAR-1995; 95JP-0094137.
 XX (SHIS) SHISEIDO CO LTD.
 XX WPI: 1996-512663/51.
 XX N-PSDB: AAT58286.
 XX
 XX DNA repair enzyme and related DNA - specifically recognises
 XX cyclobutane-type dimer and (6-4)-bound product induced by exposure
 XX to UV light
 XX

Qy 427 VKARVLENDVSVKSDVLDLPLCOELNIPVLVDHNNIV--PGLRSGSLDM--PLIP 482
 Db 399 cknrlvleddvgwtvhdllpveceelnipmvlvdyhhnncfcpahiregtldisdplqe 458
 Qy 483 TIRETWTRKGTOKOHVSESADPTAIGMKRRRAHSRVDFFPDCDPTMDLMTAEKEQA 542
 Db 459 riantwkrkgikgmhysepcd-gavtprdrkrhrprvmtlppcpdmdlmeakdkega 517
 Qy 543 VFELCRRYEL-----QNPPC-----PLEIM 562
 Db 518 vfelmtfklpgfekindmvpdyrdndenpappvkapkkkgkgrkrttdeaaapeeve 577
 Qy 563 GPEYDQTRDG-----YYPGAEKRLTARKRRSRK-----EEVEED 597
 Db 578 vpeeeramgpyrnrvywpplgceewlkpkrevkkgvpeevede 621

RESULT 6
 AAY44503
 ID AAY44503 standard; Protein: 626 AA.
 AC AAY44503;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE B. subtilis delta228-UV damage endonuclease.
 KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
 KW glutathione-S-transferase signal peptide; uvex+ gene product;
 KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
 KW abasic site; platinum diaduct; mismatched nucleotide pairing;
 KW nucleotide alkylation; skin cancer.
 XX
 OS Bacillus subtilis.
 XX
 PN WO9963828-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX
 PA (UYEM-) UNIV EMORY.
 XX
 PI Doetsch PW, Kaur B, Avery AM;
 XX
 DR WPI; 2000-116417/10.
 XX
 PT A new truncated ultraviolet damage endonuclease for treatment of skin
 PT cancers -
 XX
 PS Claim 16; Page 59; 133pp; English.
 XX

CC The present sequence is delta228-UV damage endonuclease from B. subtilis.
 CC Delta228-UVDE is a 288 amino acid deletion of the N-terminal of the
 CC uvex+ gene product. This is expressed in frame with a GST leader sequence
 CC to generate a fusion protein. This provides stable endonuclease fragments
 CC for cleaving a double-stranded DNA molecule that has a distorted
 CC structure resulting from UV radiation damage, a photoproduct, an abasic
 CC site, mismatched nucleotide pairing, platinum diaduct, an intercalated
 CC molecule or alkylation of a nucleotide. Uvexl can be used in compositions
 CC for internal or topical application and as a therapeutic agent for skin
 CC cancers.
 XX
 SQ Sequence 626 AA;

Query Match 30.9%; Score 961.5; DB 21; Length 626;
 Best Local Similarity 36.2%; Pred. No. 5.9e-70;
 Matches 233; Conservative 85; Mismatches 145; Indels 181; Gaps 18;

Qy 110 ARSSKKKATSDSVSKIDESVASYDSST-----HLRRSR-----SKKPVNYSSESE- 157
 Db 3 srkskaaaltdtpqsesstfsstldssapsparnilrsgnllqpsssekdrdhekrsgel 62
 Qy 158 -----SPEQISKATKKVKOKKEEYVEVDEK 184
 Db 63 agrmmgkdanghclregkeqegvkmaleglarmerrlqratkrqkqgleed----- 114
 Qy 185 SUKNSSSEDEFVPEQLETPISKRRS-----RSSAKNLEKESTMLNDHHA----- 232
 Db 115 gippsvsvrf-ptapyhhkstnaeacekpvkthskdvereaeigvddvkvmeapaat 173
 Qy 233 -----PREMPDCLDK-----PIPWGRILGYACLTILRSMKERVPCSTCRIT 275
 Db 174 nliepedaqdaaagaarppavvnssylplpwkgrlgyaclntylrnakppifssstcrma 233
 Qy 276 TI-----QRDGHESVKOLGTQNVLDLILVNNHNF 307
 Db 234 sivdhrhplqfedephehlknkpkdskkepdelghkfvqelglanardivkmclw----- 288
 Qy 308 IHFMKVSSDLFPFASHAKYGYTL-EFAOSHLEEVGKLANKYNNHRLTMHPGQYQTASPRE 366
 Db 289 -----fpfashpvhgkyklapfasevtaeagrvaaelghritthpgqftqlgsprk 338
 Qy 367 VVDSAI RDLAYHDEILSRMKLINEOLNKDAVLI IHLGGTFECKKETLDRFRKNYORLSDS 426
 Db 339 evesaairdleyhdellisliklpedqndavmlihmvgqfkdkaatlerfkrnyarlsgs 398
 Qy 427 VKARVLENDVSVKSDVLDLPLCQELNIPVLVDHNNIV--PGLRSGSLDM--PLIP 482
 Db 399 cknrlvleddvgwtvhdllpveceelnipmvlvdyhhnncfcpahiregtldisdplqe 458
 Qy 483 TIRETWTRKGTOKOHVSESADPTAIGMKRRRAHSRVDFFPDCDPTMDLMTAEKEQA 542
 Db 459 riantwkrkgikgmhysepcd-gavtprdrkrhrprvmtlppcpdmdlmeakdkega 517
 Qy 543 VFELCRRYEL-----QNPPC-----PLEIM 562
 Db 518 vfelmtfklpgfekindmvpdyrdndenpappvkapkkkgkgrkrttdeaaapeeve 577
 Qy 563 GPEYDQTRDG-----YYPGAEKRLTARKRRSRK-----EEVEED 597
 Db 578 vpeeeramgpyrnrvywpplgceewlkpkrevkkgvpeevede 621
 XX
 AC AAY44505;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE D. radiodurans delta228-UV damage endonuclease.
 XX
 KW Delta228-UVDE; ultraviolet damage endonuclease; GST signal peptide;
 KW glutathione-S-transferase signal peptide; uvex+ gene product;
 KW UV irradiation; DNA damage; UV radiation damage; photoproduct;
 KW abasic site; platinum diaduct; mismatched nucleotide pairing;
 KW nucleotide alkylation; skin cancer.
 XX
 OS Dienococcus radiodurans.
 XX
 PN WO9963828-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 08-JUN-1999; 99WO-US12910.
 XX
 PR 08-JUN-1998; 98US-0088521.
 PR 18-MAY-1999; 99US-0134752.
 XX


```
Qy 556 PCPLEIMGEYDQTRDGYYPGAEKRLTARKRRSRKEVEE 596
      | : | | | | | | | | | | | | | | | | | |
Db 717 ---lslnaeskipnd----ksansklatrlmkkfkæird 750
```

RESULT	9
AAB95546	
ID	AAB95546 standard; Protein; 612 AA.
XX	
AC	AAB95546;

Query Match 4.5%; Score 140.5; DB 22; Length 612;
Best Local Similarity 21.2%; Pred. No. 0.0072;

Matches	100;	Conservative	78;	Mismatches	150;	Indels	143;	Gaps
Qy	48	DTLKSLPHSSKTTISMLPQVNIAGNSFSAETPVDLKKNETELANI	SGPHKKSTSTSTR	107				
Dd	108	eatknveiasrntseqdga-----gtemrvklqeenekl-----	grseele	152				
Qy	108	KRARSSKKYATTSVD--KIDESVASYDSS-----THLRSSRSKXPPVNYN	SSSESE	157				
Dd	153	rvaqlqrgldlkdgdeakaketikkyeqelrqleaalvnhvk--eekeavsa	rralene	210				
Qy	158	SEE--QISKATNKVKQ-----KEEEYVEEVDKSLKNSSSDPEPVPQE	LTPISK	209				
Dd	211	leaaglnsqttqeqkalsekseeqeql--rrlknemenew-----hlgtiek	262					
Qy	210	RRRSRSNAKLEKSTM----NLDDHAPREMFCDLDPWPWRGRUGYACLN	TILRSMKER	265				
Dd	263	iqkemadiveaststlelpqldideykeknre-----laemqrqlkek	306					
Qy	266	VFCSTRITTIQ--RGLESVK-----QLGTGVLDLLIKLEVNNHNF	GIHFMRVSDDL	318				
Dd	307	tlaeeksrltamknqgmrlmeeelrdyqragdeaitkrqlle-----qtikdl-	355					
Qy	319	PFASHAKGYTLEFAOSHLEE-----VGKLANKYN-----HRLTMHPG	356					
Dd	356	-----eyele-akshlkddrsrlvqmedkvsglemeleerennsdllterisrse	406					
Qy	357	QYTQIASPREVVVDSATRDAYHDEILSRMKLNEQLNKO-AVLIIHLGGTFEGK	KETL--	413				
Dd	407	qmegltn--ellqeraarqldcedksl-----erqpkdklsriihlegsy	rsskeglv	459				
Qy	414	-----DRPKNTQ-----RLSDSVKARLVLENDVDSWOD	444					
Dd	460	qmeariaeledrieseedranqlsnrlkerkvk-elvmqvvddehlstd	509					
RESULT	10							
AAM60785								
ID	AAM60785	standard; protein:	733 AA.					
XX	AAM60785;							
AC								
XX								
DT	08-SEP-1998	(first entry)						
XX								
DE		Deletion mutant delta481-520 of Elongin A.						
XX								
KW	Elongin A:	rat; elongation activity; stimulation; RNA polymerase II;						
KW	binding; Elongin B; Elongin C; regulation; transcription activity;							
KW	RNA polymerase; in vitro transcriptional assay; deletion mutant.							
XX	Synthetic.							
OS	Rattus sp.							
XX								
PN	WC9814574-A1.							
XX								
PD	09-APR-1998.							
XX								
PF	02-OCT-1997;	97WO-US17992.						
XX								
PR	04-OCT-1996;	96US-0725459.						
XX								
PA	(OKLA-) OKLAHOMA MEDICAL RES FOUND.							
XX								
PI	Conaway JW,	Conaway RC;						
XX								
DR	WPI; 1998-286420/25.							
XX								
PT	Isolated Elongin A and C and fragments - used for regulating							
PT	transcriptional activation of RNA polymerase, particularly for use							
PT	in vitro assays and systems							
XX								
PS	disclosure; Pages 80-83; 185pp; English.							
XX								
XC	AAM60783-92 represent deletion mutants created to define Elongin A							


```

Db 366 -----veekleetygyfrefkfttrilvkrkrnkeqkklkedekkli 408
Qy 509 SG-----MKRRAHSDRVF-----DFPPCDPTMDLMIEAKKEQAVFELCRRYEL 552
Db 409 aaeepddekklkdsdkvvvpvknkssfpd-----kfrapdkkrmtf-----yrl 456
Qy 553 QN--PPCP-----LEIMGPEVDQTRDGYPPGAEKRLTA-----RKRRSRKEEVEDE 598
Db 457 selfpiivrkdnelavsgdcmdskvng-----kklkstfnfkrrnrklkerkmqe 507

RESULT 12
AAW54145
ID AAW54145 standard; Protein; 1639 AA.
AC AAW54145;
XX
DT 23-SEP-1998 (first entry)
XX
DE P. falciparum synthetic gp190 protein.
XX
KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW monoclonal antibody; passive immunisation; parasite.
XX
OS Plasmodium falciparum.
OS Synthetic.
XX
PN WO9814583-A2.
XX
PD 09-APR-1998.
XX
PF 02-OCT-1997; 97WO-EP05441.
XX
PR 02-OCT-1996; 96DE-4040817.
XX
PA (BUJA/) BUJARD H.
XX
PI Bujard H, Pan W, Tolle R;
XX
DR WPI; 1998-240088/21.
DR N-PSDB; AAV21451, AAV35363.
XX
PT Recombinant production of complete gp190/MSP-1 Plasmodium surface
PT protein - useful in anti-malaria vaccines, also stabilising genes by
PT reducing their AT content
XX
PS Example 1; Fig 3c; 48pp; German.
XX
CC This sequence represents a modified Plasmodium falciparum gp190/MSP-1
CC (merozoite surface) protein. The gene encoding this protein has been
CC stabilised by reducing the AT content of the nucleotide sequence. Such a
CC protein is useful in vaccines against malaria or for producing monoclonal
CC antibodies (for passive immunisation). The complete gp190 protein can now
CC be produced outside the parasite and has, at least over extended regions,
CC the native pattern of folding. Larger amounts of the protein can be
CC produced recombinantly than would be possible using the parasites as
CC source.
XX
SQ Sequence 1639 AA;

Query Match 4.4%; Score 136.5; DB 19; Length 1639;
Best Local Similarity 19.5%; Pred. No. 0.066;
Matches 130; Conservative 107; Mismatches 212; Indels 217; Gaps 35;

Qy 17 FTILKQKAFKNHPCVPSVCHTYS---RFHCLPDTLKSLLPMSKTTLSMLPQVNIGAN 73
Db 185 follrakl---ndvcandycqipnlkiranell-dvllkklv----- 221
Qy 74 SFSAEPTPVDLKKENETELANTISGPHKSTSTSTRKRARSKKATDSVSKIDESVASYD 133
Db 222 -fgyrkpldnldkdvngmedyl---kkn-----kktieninelieeskktd 264

```

```

Qy 134 SSTHLRRSSRSKKPVN-----YNSSESESEEOISKATKKVKOKEEEXVEEYDEK-- 184
Db 265 kknatkkeekkklyqaydlsyknqle-eahmlisvlekrldtlkknenkellidkn 323
Qy 185 SLKN-----ESSODEFEPVPEOLET-----PIKRRRRSRSSA 217
Db 324 eiknppansgntpntllidknkkieehekeiaktikfnidsiftqpleyleylrekn 383
Qy 218 KULE-----KESTMNLDDHAPREMFCCLDKPIPWGRGUYACLNITLRSM----- 262
Db 384 knidisakvetkest-----epneyngvtyp-----lsyndnnaaIneInsfgdlin 431
Qy 263 -----KERVFCSTCRITTIQDGLSVKQL---GTQNVLDLILVE-- 301
Db 432 pfdytkepsknlytdnerkkfineikekkiesdkksyedrsksinditkeyekl 491
Qy 302 -----WNHFGIHFMRVSSDLFPFAS--HAKYCYTLEFAQSHLEEVGKILANKYNHRL 351
Db 492 lneiydskfnnni-----dltnfekmmgkrysykveklthntfasyenskhniek 542
Qy 352 TMHPCGYTOIASPREVVVDSAIRDLAYHDEILSRMK-----LNFOINKDAVLIHILGGTF 406
Db 543 ltkalkymedyslrnlivve---kelkyyknlskieneietlivenikkdeeql-----f 593
Qy 407 EGKKEPTLDRFRKNYQ--RLSDSVKAR---LVLENDVSVSVQDILLPLCOELNIPILVLDW 460
Db 594 e-kkltkdenkpkdekllvdsdvkvqkvllmnikdelkkktqllik-----nvel--- 643
Qy 461 HHHNI--VPGTLREGS-----LD---LMPLTIPTRETWTRKGTQKHYSSES 502
Db 644 -knhivpnsyqenkeqpyylivlkkeldkikvfmkpveslineekkniklegq--sdn 700
Qy 503 ADPTA---ISGMKRRRAHSDRVDFPPCDPTMDLMEAKEKEQAVFELCRRYELQNPQPL 559
Db 701 sepstegeitgattkpgqagqsaalegd---svqagaqeqkda-----qpypvpv 746
Qy 560 EIMGPE 565
Db 747 pv--pe 750

RESULT 13
AAW5657
ID AAY85657 standard; Protein; 1341 AA.
XX
AC AAY85657;
XX
DT 12-FEB-2001 (first entry)
XX
DE Human Acinus L protein sequence.
XX
KW Chromatin aggregation induction; human; Acinus; AIDS; herpes; diabetes;
KW Alzheimer's disease; Parkinson's disease; cardiac infarction;
KW brain infarction; adenovirus infection; viral hepatitis.
XX
OS Homo sapiens.
XX
PN WO200061743-A1.
XX
PD 19-OCT-2000.
XX
PF 07-APR-2000; 2000WO-JP02254.
XX
PR 09-APR-1999; 99JP-0103317.
XX
PA (SHIO ) SHIONOGI & CO LTD.
XX
PI Sahara S, Equichi Y, Tsujimoto Y;
XX
DR WPI; 2000-665130/64.
DR N-PSDB; AAC61196.
XX

```


Db 347 grtgdpiprakdkvnnlkagqkvrtndrkspslpkveemdmdddefeqptmsfesy 406
Qy 238 DCLDKPIPMWRGLGYACNLITLSMRKERVFCSTCRITTIQDGLSVKOLGT-QNVLDL 296
Db 407 lsydqp-----rkkkkvvktsg---talgekgllkkdkstknlnsa 447
Qy 297 IKLVEWHNFGIHPM-----RVSSDLFPF-----ASHAKYG--YTLEFAQS-- 335
Db 448 qklpanenksdklpagaeptprkvptdvlpalpdlipaiqtnyrplpslelissfq 507
Qy 336 -----HLEEVGKLANKYNHRLTMHPGQYQTIASPREVVVDSAIRDLAYHDEILSR 385
Db 508 pkrkafspqeeeaagftgrmismkmqvygskcaylpmmtlhqccirvlnknnidsiq 567
Qy 386 MKNLEQLN-----KDAVLIHILGGTF-EGKKTLDLRFKKNYORLSDSVKARLVLENDV 438
Db 568 lyrieecnhvlietdqlwkvchrdfkeerpeeyeswremyirlgdareqirlltnni 627

RESULT 15

AAW60787
ID AAW60787 standard; protein; 753 AA.

AC AAW60787;

XX 08-SEP-1998 (first entry)

XX Deletion mutant delta546-565 of Elongin A.

XX Elongin A; rat; elongation activity; stimulation; RNA polymerase II;
XX binding; Elongin B; Elongin C; regulation; transcription activity;
XX RNA polymerase; in vitro transcriptional assay; deletion mutant.

XX Synthetic.

OS Rattus sp.

XX W09814574-A1.

XX 09-APR-1998.

XX 02-OCT-1997; 97WO-US17992.

XX 04-OCT-1996; 96US-0725459.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX Conaway JW, Conaway RC;

XX WPI; 1998-286420/25.

XX Isolated Elongin A and C and fragments - used for regulating
XX transcriptional activation of RNA polymerase, particularly for use
XX in in vitro assays and systems

PS Disclosure; Pages 86-88; 185pp; English.

XX AAW60783-92 represent deletion mutants created to define the Elongin A
XX sequences required for binding to Elongins B and C and for
XX transcriptional activity. Elongin A stimulates the elongation activity
XX of RNA polymerase II. Mutants that span the transcriptional activation
XX domain from residues 400 to 730 (AAW60775) were constructed and
XX expressed in Escherichia coli. The present deletion mutant lacks amino
XX acids 546 to 565 of the wild type Elongin A (AAW60771). The Elongin
XX polypeptides and fragments can be used for regulating the transcriptional
XX activity of RNA polymerase. They can be used in in vitro transcriptional
XX assays or systems.

XX Sequence 753 AA;

Query Match 4.3%; Score 134.5; DB 19; Length 753;
Best Local Similarity 20.3%; Pred. No. 0.03;

Matches 86; Conservative 73; Mismatches 169; Indels 95; Gaps 18;
Qy 73 NSFSAETPVDLKKENETELANISGPHKKSTSTSTKRARSSKKKATDSVSKIDSVASY 132
Db 243 sshkekrrpvdargdekssvmgrekshkasskeesrll-----sedsakeklpsvsvkk 296
Qy 133 DSSTHLRRSSRSRKKPVN--YNSSESSESEPEQISKATKKVKQKEEYVEEVEDEKSLKNES 190
Db 297 ekd---reunsikkklspaldvasdnhfkpkhkdksekikskdnkqsvdsvd-----s 346
Qy 191 SSDEFEPV-----VPEQLETPISKRRRS--RSSAKNLEKESTMNLDDHAPR-----EMF 237
Db 347 grtgdpiprakdkvnnlkagqkvrtndrkspslpkveemdmdddefeqptmsfesy 406
Qy 238 DCLDKPIPMWRGLGYACNLITLSMRKERVFCSTCRITTIQDGLSVKOLGT-QNVLDL 296
Db 407 lsydqp-----rkkkkvvktsg---talgekgllkkdkstknlnsa 447
Qy 297 IKLVEWHNFGIHPM-----RVSSDLFPF-----ASHAKYG--YTLEFAQS-- 335
Db 448 qklpanenksdklpagaeptprkvptdvlpalpdlipaiqtnyrplpslelissfq 507
Qy 336 -----HLEEVGKLANKYNHRLTMHPGQYQTIASPREVVVDSAIRDL 376
Db 508 pkrkafspqeeeaagftgrmismkmqvygskcaylptfegvgvpyvslpvlertcpdq 567
Qy 377 AYHDEILSKMKLNEQLNKDAVLIHILGGTF-EGKKTLDLRFKKNYORLSDSVKARLVLEN 435
Db 568 lyrieecnhvlietdqlwkvchrdfkeerpeeyeswremyirlgdareqirllit 624
Qy 436 DDV 438
Db 625 nni 627

Search completed: January 15, 2002, 13:49:59
Job time: 288 sec

